



KPC Groep

Lezen op AVI-niveau

De schaling, normering en validering
van de nieuwe AVI-toetskaarten

Jacqueline Visser

Lezen op AVI-niveau

Lezen op AVI-niveau

De schaling, normering en validering van de nieuwe AVI-toetskaarten

**Een wetenschappelijke proeve op het gebied van
de Sociale Wetenschappen**

Proefschrift

**ter verkrijging van de graad van doctor
aan de Katholieke Universiteit Nijmegen,
volgens besluit van het College van Decanen
in het openbaar te verdedigen
op vrijdag 20 juni 1997 des namiddags om 1.30 uur precies**

door

Jacqueline Johanna Cornelia Maria Visser

**geboren op 2 september 1962
te Bergh**

Promotor:	Prof. dr. R.M. van den Berg
Copromotores:	Dr. M.J.M. Voeten
	Dr. C.A.J. Aarnoutse

Leden manuscriptcommissie:
Prof. L.T.W. Verhoeven (voorzitter)
Prof. A.J.W.M. Thomassen
Dr. M.G.H. Jansen (Rijksuniversiteit Groningen)

Voor Lynn

Visser, Jacqueline Johanna Cornelia Maria

Lezen op AVI-niveau: de schaling, normering en validering van de nieuwe AVI-toetskaarten / Jacqueline Johanna Cornelia Maria Visser. - 's-Hertogenbosch: KPC Groep. - Proefschrift Nijmegen. - Met lit. opgave. - Met samenvatting in het Engels. ISBN 90-6755-118-X

Trefw.: Technisch lezen ; basisonderwijs / toetsontwikkeling / ontwikkeling leesvaardigheid.

© Copyright: KPC Groep 's-Hertogenbosch

VOORWOORD

Sinds het verschijnen van de nieuwe AVI-toetskaarten in maart 1994 is mij vele malen de vraag gesteld: de toetskaarten zien er mooi uit, maar hoe zit het met de verantwoording van de AVI-toetskaarten, waar is dat beschreven? Het antwoord dat ikzelf telkens gaf was: daar wordt hard aan gewerkt. Vanaf nu luidt het antwoord: daar heb ik een proefschrift over geschreven. Met dit proefschrift heb ik duidelijk willen maken wat de waarde van de AVI-toetskaarten is voor het leesonderwijs. Tevens wil ik het belang aangeven van het verwerven van een goede technische leesvaardigheid voor het kunnen functioneren in onze maatschappij, waarin het gedrukte woord zo belangrijk is. Het proefschrift draag ik op aan mijn petekind Lynn, en via haar aan alle kinderen die nog aan het begin staan van het verwerven van de leesvaardigheid.

Bij het uitvoeren van het onderzoek, maar ook tijdens het schrijven van het proefschrift, heb ik op velen mogen rekenen. Allereerst wil ik de leerlingen en leerkrachten van de scholen bedanken die hebben deelgenomen aan de twee try-outs en het uiteindelijke onderzoek. De studenten en docenten van de Pabo's in Breda, Eindhoven, Helmond, Hengelo, Utrecht, Rotterdam en Zwolle ik bedanken voor hun hulp bij het afnemen van de AVI-toetskaarten.

Graag wil ik mijn promotor Dolf van den Berg en de copromotoren Rinus Voeten en Cor Aarnoutse bedanken voor hun begeleiding. Zowel Dolf en Jan Arts als de directie van KPC Groep, wil ik bedanken voor de mogelijkheden die zij me geboden hebben om te werken aan mijn promotie-onderzoek.

Zonder mijn directe collega's van KPC Groep zou het onderzoek en het schrijven van dit proefschrift een stuk moeilijker zijn geweest. Speciaal wil ik noemen mijn mede-AVI-consultante Anne ter Beek die onmisbaar was bij de uitvoering van het onderzoek op de scholen, en Cecile Winnubst en Jos Theunissen die mij met hun bemoedigende woorden hebben gesteund wanneer het proefschrift mij tot wanhoop bracht.

Diny Albers en Lieke Eijsackers van de Bidoc hebben zeer geduldig al mijn literatuuravragen verwerkt. Marja Nieuwenhuis en Ton van Son van de afdeling Vormgeving adviseerden me over de lay-out van mijn proefschrift. Annique Knotter verzorgde het omslagontwerp. Graag wil ik hen voor hun ondersteuning bedanken.

Jan van Leeuwe en Nicole van Kan hebben ervoor gezorgd dat ik over de juiste data beschikte. Margo Jansen van de Rijksuniversiteit Groningen leverde een belangrijke bijdrage aan de schaling van de toetskaarten. Lee Ann Weeks van Scientific Editing Service zorgde voor de Engelse vertaling van de samenvatting. Lieke de Jong van de vakgroep Arabisch voorzag me van het Arabische woord in de inleiding. Ook hen wil ik graag bedanken.

Als laatste wil ik Erik bedanken voor zijn steun en adviezen bij het schrijven van dit proefschrift, dat nogal eens ten koste ging van onze gezamenlijke vrije tijd.

Jacqueline Visser
Dieren, mei 1997

INHOUDSOPGAVE

VOORWOORD	vii
INLEIDING	1
1 AVI-TOETSKAARTEN: TECHNISCH LEZEN GEMETEN	5
1.1 De AVI-toetskaarten	5
1.1.1 Historisch perspectief	5
1.1.2 Een beschrijving van de AVI-toetskaarten	9
1.1.3 Vernieuwing van de AVI-toetskaarten	11
1.2 Technisch lezen	12
1.2.1 De eenheid van verwerking bij woordherkenning	14
1.2.2 De rol van fonologische processen	16
1.2.3 De rol van de context	19
1.2.4 Woordherkenning als een automatisch proces	21
1.3 Hoe verwerven kinderen de technische leesvaardigheid?	24
1.3.1 Modellen voor de verwerving van de technische leesvaardigheid	24
1.3.2 Onderwijs in technisch lezen	28
1.4 Technisch lezen, begrijpend lezen en woordenschat	35
1.4.1 Technisch lezen en begrijpend lezen	35
1.4.2 De relatie van technisch en begrijpend lezen met woordenschat	38
1.5 Het meten van technisch lezen	40
1.5.1 Eisen voor een toets voor technisch lezen	41
1.5.2 De AVI-toets vergeleken met andere toetsen voor technisch lezen	43
2 DE ONTWIKKELING VAN DE NIEUWE AVI-TOETSKAARTEN	51
2.1 De criteria voor de negen AVI-niveaus	51
2.1.1 De leesbaarheid van teksten: leesbaarheidsformules	51
2.1.2 Leesindex A en de negen AVI-niveaus	56
2.2 De constructie van de toetskaarten	61
2.2.1 Nieuwe teksten	62
2.2.2 De eerste try-out	63
2.2.3 Beoordeling door leerkrachten	63
2.2.4 De tweede try-out	65
2.3 De onderzoeksvragen	71
3 SCHALING EN NORMERING	73
3.1 De onderzoeksofzet	73
3.1.1 De steekproef	73
3.1.2 De dataverzameling	74
3.2 Het bepalen van de grenswaarden	76
3.2.1 Uitgangspunten	76
3.2.2 Schaling op basis van de leestijden	78
3.2.3 Grenswaarden op basis van de leestijden	87
3.2.4 Correctie voor leesfouten	91
3.2.5 De uiteindelijke grenswaarden	95

3.3	De betrouwbaarheid van de AVI-toetskaarten	96
3.4	De normering	105
3.5	Samenvatting en conclusies	110
4	VALIDITEITSONDERZOEK	113
4.1	De onderzoeksopzet	113
4.1.1	De steekproef	113
4.1.2	Analyseprocedure	125
4.2	De relatie tussen AVI en EMT	128
4.2.1	Het basismodel	128
4.2.2	Parallelliteit van A- en B-versies van AVI en EMT	131
4.2.3	De samenhang tussen AVI en EMT	137
4.3	De relatie met begrijpend lezen en woordenschat per cohort per periode	139
4.3.1	Het basismodel	139
4.3.2	De relatie tussen Begrijpend Lezen en Woordenschat	140
4.3.3	De relatie tussen AVI en Begrijpend Lezen en Woordenschat	142
4.4	De relatie met begrijpend lezen en woordenschat: longitudinaal	144
4.4.1	De basismodellen	144
4.4.2	De longitudinale relatie tussen AVI en Begrijpend Lezen	147
4.4.3	De longitudinale relatie tussen AVI en Woordenschat	151
4.4.4	Interpretatie van de uiteindelijke modellen	156
4.5	Conclusies	160
5	DE KWALITEIT VAN DE NIEUWE AVI-TOETSKAARTEN	163
5.1	Samenvatting en conclusies	163
5.1.1	De AVI-toetskaarten	163
5.1.2	Herziening van de AVI-toetskaarten	165
5.1.3	De schaling en normering	166
5.1.4	De validering	169
5.2	Discussie	171
5.2.1	De psychometrische kwaliteit	171
5.2.2	De waarde voor de onderwijspraktijk	173
5.2.3	Bijdrage aan theorievorming en suggesties voor nader onderzoek	176
	SUMMARY	181
	LITERATUUR	187

BIJLAGEN

1	Overzicht van de AVI-niveaus	203
2	Voorbeelden nieuwe toetskaarten	205
3	Oorspronkelijke indeling AVI-niveaus	207
4	Korte instructie try-out januari 1992	209
5	Voorbeeld score-formulier	211
6	Descriptieve statistieken per toetskaart in de try-out januari 1992	213
7	Korte instructie normeringsonderzoek	215
8	Omzettingstabellen ten behoeve van toewijzen toetskaarten	217
9	Descriptieve statistieken per toetskaart in het normeringsonderzoek	219
10	Uitgeteste grenswaarden voor fouten	225
11	Gebruikte toetsen voor technisch lezen, begrijpend lezen en woordenschat	227
12	Descriptieve statistieken van de toetsen voor technisch lezen, begrijpend lezen en woordenschat in het oorspronkelijke bestand	229
13	Spreidingsdiagrammen van scores op de AVI-toetskaarten met scores op de EMT	233
14	Ongestandaardiseerde parameterschattingen in het uiteindelijke model voor parallelliteit AVI en EMT	239
15	Ongestandaardiseerde parameterschattingen in het uiteindelijke model voor parallelliteit AVI en EMT bij simultane analyses voor twee cohorten	243
16	Ongestandaardiseerde parameterschattingen in het uiteindelijke model per periode voor de relatie van AVI met begrijpend lezen en woordenschat	247
17	Ongestandaardiseerde parameterschattingen in de uiteindelijke longitudinale modellen voor de relatie van AVI met begrijpend lezen	261
18	Ongestandaardiseerde parameterschattingen in de uiteindelijke longitudinale modellen voor de relatie van AVI met woordenschat	271
19	Correlatiematrices, gemiddelden en standaardafwijkingen voor Lisrel-analyses	281
CURRICULUM VITAE		293

INLEIDING

Stelt u zich voor dat u het volgende Arabische woord moet voorlezen:

قَرَأَ

Om dit woord te kunnen lezen moet u een aantal dingen weten. Zo dient u van rechts naar links te lezen. Daarnaast zijn er bepaalde regels voor het omzetten van de lettertekens in klanken. Voor veel klanken zijn aparte tekens beschikbaar, maar soms kan eenzelfde teken voor verschillende klanken staan, of kan eenzelfde klank door meerdere tekens worden weergegeven. Het ontsleutelen van geschreven informatie noemen we technisch lezen. We zullen u niet vermoeien met het decoderen van bovenstaande Arabische tekens¹, maar duidelijk is dat het leren ontsleutelen van geschreven of gedrukte woorden een van de moeilijkste opgaven is waarvoor een beginnende lezer staat. Het technisch kunnen lezen is ook een belangrijke voorwaarde voor het uiteindelijke doel van lezen: begrijpen wat er geschreven staat.

Er bestaan grote verschillen tussen leerlingen in de mate waarin de technische leesvaardigheid geautomatiseerd is. Om de ontwikkeling van deze vaardigheid goed te kunnen vaststellen, zijn instrumenten, toetsen nodig. Een bekend instrument voor het meten van de technische leesvaardigheid is de AVI-toets. In maart 1994 is een nieuwe, geheel herziene versie van deze toets verschenen, zo'n 20 jaar nadat het KPC de oorspronkelijke uitbracht.

De AVI-toetskaarten, de AVI-niveaus en de AVI-niveaulijsten zijn bekende begrippen in het basisonderwijs. Maar wat is de oorsprong daarvan, waar staan de letters A V I voor? AVI staat voor 'Analyse Van Individualiseringsvormen', een innovatieproject van het KPC uit 1972. Doel van het project was om hulpmiddelen en werkwijzen te ontwikkelen voor de individualisering van het gehele leesonderwijs in de basisschool, zowel voor de lestechniek, het leesbegrip als de leesbeleving. Vooral de uitwerkingen voor het technisch leesonderwijs zijn in de praktijk ingevoerd en worden nog steeds gebruikt. Dit betreft de AVI-toetskaarten en de AVI-niveaulijsten voor het technisch lezen.

Voor het onderwijs in technisch lezen zijn negen AVI-niveaus gecreëerd (Van den Berg & Te Lintelo, 1977b). Deze niveaus worden enerzijds gebruikt om leesboeken en teksten in te delen naar moeilijkheidsgraad. Daarvoor zijn AVI-niveaulijsten beschikbaar. Anderzijds wordt de technische leesvaardigheid van de leerlingen ingedeeld naar deze negen niveaus, waarvoor de AVI-toetskaarten zijn ontwikkeld. Het doel van de AVI-toetskaarten is het vaststellen van het niveau van technisch lezen van de leerlingen om zo, ten behoeve van het onderwijs in technisch lezen, leesstof te kunnen kiezen die aansluit bij dat niveau.

¹ Het afgebeelde woord wordt uitgesproken als "qara'a" en betekent "lezen".

In het proefschrift beschrijven we de ontwikkeling van de nieuwe AVI-toetskaarten en de schaling, normering en validering van de toets. De vraag die we willen beantwoorden is: hoe betrouwbaar en valide zijn de nieuwe AVI-toetskaarten om de technische leesvaardigheid van leerlingen vast te stellen? We onderzoeken de relatie tussen de moeilijkheidsgraad van de toetskaarten en de technische leesvaardigheid van de leerlingen.

Het proefschrift is als volgt opgebouwd.

In hoofdstuk 1 beschrijven we eerst de AVI-toetskaarten: de oorsprong ervan, hoe de toets eruit ziet en waarom besloten is nieuwe toetskaarten te ontwikkelen. Vervolgens komen we, vanuit de beschrijving van de huidige kennis en theorieën over het proces van technisch lezen, de verwerving van de technische leesvaardigheid en de relatie van technisch lezen met begrijpend lezen en woordenschat, tot een formulering van eisen waaraan een toets voor technisch lezen dient te voldoen. Een toets bedoeld om vast te stellen in welke mate de technische leesvaardigheid geautomatiseerd is, dient te bestaan uit woorden, zinnen of teksten van uiteenlopende moeilijkheidsgraad en moet zowel de nauwkeurigheid als de snelheid van woordherkenning meten. We zullen de AVI-toetskaarten tevens vergelijken met andere toetsen bedoeld voor het meten van de technische leesvaardigheid.

In hoofdstuk 2 beschrijven we de wijze waarop de nieuwe AVI-toetskaarten zijn geconstrueerd. We bespreken eerst de criteria voor de negen AVI-niveaus: waarop ze zijn gebaseerd en hoe ze zijn aangepast voor de nieuwe AVI-toetskaarten. Daarbij maken we een uitstapje naar leesbaarheidsformules en de Leesindex A van Brouwer in het bijzonder, aangezien de indeling van de niveaus 5 tot en met 9 daarop gebaseerd is. Vervolgens wordt de constructie van de nieuwe AVI-toetskaarten beschreven, die in een aantal fasen is verlopen. Aan het eind van dit hoofdstuk presenteren we de zes onderzoeksvragen die centraal staan in het onderzoek dat in dit proefschrift wordt beschreven.

In hoofdstuk 3 staat het onderzoek naar de schaling en normering van de toetskaarten centraal. Uitgaande van een 'latente trek'-model van de leestijden proberen we de toetskaarten te schalen op basis van empirische gegevens. Daarbij gaan we tevens na of de toetskaarten daadwerkelijk oplopen in moeilijkheidsgraad en of deze geschatte moeilijkheidsgraden overeenkomen met de veronderstelde moeilijkheidsgraden. Vervolgens bepalen we de grenswaarden voor elke toetskaart, dat wil zeggen het aantal fouten en de leestijd die aangeven of een leerling het AVI-niveau beheerst waarvoor de toetskaart bedoeld is. Tot slot onderzoeken we de betrouwbaarheid van de AVI-toets en presenteren we de normeringsgegevens van de AVI-toets voor een aantal tijdstippen in het basisonderwijs.

Centraal in hoofdstuk 4 staat de validering van de AVI-toets. Met behulp van lineair structurele (Lisrel)modellen gaan we eerst na of de A- en B-versie van de AVI-toets als paralleltests beschouwd mogen worden. Vervolgens onderzoeken we met dezelfde Lisrelmodellen de relatie tussen de AVI-toetskaarten en de Eén-Minuut-Test, de andere toets voor technisch lezen die in het onderzoek is afgenomen. We gaan na of deze toetsen hetzelfde meten (convergente validiteit). Tot slot onderzoeken we de relatie tussen de

AVI-toetskaarten en toetsen voor begrijpend lezen en woordenschat (divergente validiteit). Daarbij zullen we eerst cross-sectioneel de correlaties schatten tussen de AVI-toets enerzijds en de toetsen voor begrijpend lezen en woordenschat anderzijds. Daarna ontwikkelen we longitudinale modellen met behulp waarvan nagegaan wordt wat de invloed is van een toename in de technische leesvaardigheid (gemeten met de AVI-toetsen) op de prestaties bij begrijpend lezen en woordenschat.

Tot slot geven we in hoofdstuk 5 eerst een samenvatting van de voorgaande hoofdstukken, waarbij we tevens de conclusies ten aanzien van de onderzoeksvragen presenteren. Vervolgens gaan we in op de psychometrische kwaliteit van de toetskaarten, de waarde van de toetskaarten en van het onderzoek voor de onderwijspraktijk. Tot slot schetsen we mogelijke wegen voor verder onderzoek. Degenen die snel een inzicht willen hebben in de opzet en de resultaten van het onderzoek kunnen zich beperken tot het lezen van dit laatste hoofdstuk.

1 AVI-TOETSKAARTEN: TECHNISCH LEZEN GEMETEN

Centraal in dit proefschrift staat de schaling, normering en validering van de nieuwe AVI-toetskaarten. Deze toetskaarten zijn bedoeld voor het meten van de technische leesvaardigheid in het basisonderwijs. Om dat technisch lezen te kunnen meten, moeten we een goed inzicht hebben in:

- de factoren die een rol spelen in het proces van technisch lezen;
 - de wijze waarop kinderen de technische leesvaardigheid verwerven en verder ontwikkelen, en
 - de relatie van technisch lezen met het uiteindelijke doel: het begrijpen van teksten.
- Op basis van deze inzichten zullen we in dit hoofdstuk bepalen welke metingen van (aspecten van) de technische leesvaardigheid van belang zijn voor het reguliere basisonderwijs. Voor het bepalen van de validiteit van de nieuwe AVI-toetskaarten zijn deze gegevens van essentieel belang.

In paragraaf 1.1 beginnen we met een schets van de historie van de AVI-toetskaarten. We geven tevens een korte beschrijving van de toets, gevolgd door de redenen waarom de toetskaarten herzien zijn. De oorspronkelijke AVI-toetskaarten stammen uit de jaren zeventig. We zullen in dit hoofdstuk proberen aan te geven hoe de AVI-toets gesitueerd kan worden in de huidige inzichten omtrent het technisch lezen.

In paragraaf 1.2 schetsen we de stand van zaken met betrekking tot onze kennis over het technisch lezen: welke factoren spelen een rol in het proces van technisch lezen. In paragraaf 1.3 staan we stil bij de verwerving en de ontwikkeling van het technisch lezen, bij de fasen die daarin te onderscheiden zijn, en de rol van het onderwijs en de AVI-toets daarin.

Het technisch lezen is uiteraard geen doel op zich, maar staat in dienst van het proces van begrijpend lezen. In paragraaf 1.4 schenken we aandacht aan de relatie tussen technisch lezen enerzijds en begrijpend lezen en woordenschat anderzijds.

Rekening houdend met de inzichten over het proces van technisch lezen, de ontwikkeling van de technische leesvaardigheid en de relatie van technisch lezen met begrijpend lezen en woordenschat gaan we in paragraaf 1.5 in op de vraag: hoe kunnen we de vaardigheid in technisch lezen meten? Daarbij bespreken we de elementen waaraan een toets voor technisch lezen zou moeten voldoen, gegeven de functie van de toets en geven we aan welke keuzes gemaakt zijn bij de constructie van de AVI-toets uitgaande van het doel van de toets. Daarbij zullen we de AVI-toets tevens vergelijken met andere toetsen voor technisch lezen.

1.1 De AVI-toetskaarten

1.1.1 Historisch perspectief

De wortels van de AVI-toetskaarten zijn gelegen in de roep om individualisering van het onderwijs in de jaren zestig en zeventig. Het leerstofjaarklassensysteem dat destijds de dominante organisatievorm van het onderwijs was, en ook nu nog is, werd in toenemende mate bekritiseerd. Het systeem berust op de aanname dat alle leerlingen van een

bepaalde leeftijd dezelfde beginsituatie hebben (De Corte, Geerligs, Lagerweij, Peters & Vandenbergh, 1981). Deze aanname bleek echter niet houdbaar, en de met deze aanname samenhangende uniformiteit van het leerproces riep diverse problemen op. Leerlingen die achter raken op de klas, moeten bij de overgang van het ene leerjaar naar het andere afhaken. Doornbos (1969) toonde aan dat dit zittenblijven niet direct een nuttig effect heeft op de leerprestaties van de leerlingen. Als oplossing gaf hij aan dat de koppeling tussen leerjaar en leerstof losgelaten moest worden, hetgeen ontkoppeling van cursusduur en leerstofbeheersingseisen en differentiatie van prestatie-eisen zou betekenen. Gevraagd werd om individualisering van het onderwijs.

Deze roep om individualisering leidde ertoe dat de drie Landelijke Pedagogische Centra op zoek gingen naar organisatievormen die versnelling en/of vertraging in de aanpak van het onderwijs mogelijk moesten maken (Van den Berg, 1977). Daartoe werd een drietal ontwikkelingen in gang gezet waarin het vraagstuk van individualisering vooral vanuit schoolorganisatorische gezichtspunten werd benaderd:




- de start van de zogenaamde Jenaplanscholen, als uitvloeisel van een experiment van de ontwikkeling van niet-klassikale scholen;
- de verdere uitbouw van het systeem van het losser klasseverband, waarin vooral aandacht was voor flexibele vorderingengroepen binnen klasseverband, en
- de ontwikkeling van de zogenaamde niveaucursusorganisatie, waarbij het klasseverband werd losgelaten (zie Van den Berg, 1977).

Binnen deze drie organisatievormen trachtte men voor het leesonderwijs (met als doelstellingen de ontwikkeling van de leesteknik, het leesbegrip en de leesbeleving) een aantal leesvormen en daarbij behorende groeperingsvormen toe te passen. De leesvormen die men onderscheidde waren: technisch lezen, expressief lezen, begrijpend lezen, studerend lezen, belangstellend/informatief lezen, boekbesprekend lezen en integrerend lezen. De groeperingsvormen betroffen het klassikaal, het groepsgewijs en het individueel lezen. In tabel 1.1 hebben we het 'Schema Individualiserend Leesonderwijs' weergegeven, waarin staat aangegeven welke leesvormen en groeperingsvormen men koos bij elk van de drie doelstellingen.

Individualisering werd in eerste instantie vooral opgevat als een schoolorganisatorisch probleem. Gezien de toenemende belangstelling voor individualisering en het groeiende inzicht dat individualisering meer is dan het wijzigen van de schoolorganisatie, besloot het KPC een innovatiestrategie op te zetten. Dit nu resulteerde in 1972 in het project 'Analyse Van Individualiseringsvormen'; AVI werd geboren.

Het individualiseringsproject was bedoeld om hulpmiddelen en werkwijzen te ontwikkelen voor de individualisering van het leesonderwijs in Nederland. Men beperkte zich overigens tot het leesonderwijs omdat dat in de praktijk dikwijls het eerste werkveld vormde om het onderwijs te individualiseren en omdat voor het leesonderwijs al instrumenten (leergangen en toetsen) beschikbaar waren om het proces van individualiseren te ondersteunen en te sturen (zie Van den Berg & Te Lintelo, 1977b).

Tabel 1.1: Schema Individualiserend Leesonderwijs (overgenomen uit Van den Berg & Te Lintelo, 1977b, p.21)

Algemene doelen	Leesvormen	Groepenvormen/ situaties
Ontwikkeling en bevordering van I leestechneek 	1. technisch aanvankelijk lezen 2. technisch voortgezet lezen	a. klassikaal b. lezen in(vorderingen)- groepen
II leesbegrip 	3. begrijpend lezen 4. studerend lezen	a klassikaal b. groepslezen volgens de parallele methode c groepslezen volgens de taakverdelende methode d forumlezen
III leesbeleving 	5. belangstellend/infor- matief lezen 6. boekbesprekend lezen 7. integrerend lezen	- individueel - groepsknngen - leesknng

In 1974 werd het innovatiepakket AVI op de markt gebracht, en in 1977 verscheen het herziene AVI-pakket, waarbij de beschikbare materialen in twee delen verschenen (ook wel mappen genoemd):

- Deel 1 'Observatie en begeleiding van individualiseringsvormen' (Van den Berg & Te Lintelo, 1977a), dat een scholingsprogramma voor begeleiders en een begeleidingsprogramma voor schoolteams bevat;
- Deel 2 'Individualisering van het leesonderwijs (Van den Berg & Te Lintelo, 1977b), waarin aandacht wordt besteed aan de wijze waarop individualisering van het leesonderwijs gerealiseerd kan worden. Zowel de leestechneek, het leesbegrip als de leesbeleving komt in deel 2 aan bod.

Voor het individualiserend leesonderwijs, gericht op de drie eerder genoemde doelstellingen, koos men drie criteria om de leesstof op een geschikte wijze te kiezen en te ordenen:

- *de doelgerichtheid:*

de leesstof moet zodanig gekozen en opgebouwd zijn dat de gestelde doelen (leestechneek, leesbegrip en leesbeleving) bereikt kunnen worden.

- *de beginsituatie:*
de leesstof moet overeenstemmen met de leesvaardigheid en de verstandelijke mogelijkheden van het kind. Daarnaast is het wenselijk rekening te houden met de individuele verschillen in belangstelling en geaardheid.
- *de inhoud van de leestekst:*
de teksten moeten qua inhoud de verschillende leerlingen aanspreken." (Van den Berg & Te Lintelo, 1977b, p. 27).

Voor de keuze van de leesstof vonden Van den Berg en Te Lintelo (1977b) het laatste criterium, de inhoud, het meest belangrijk. Voor het ordenen van de leesstof naar moeilijkheidsgraad kozen ze voor leesbaarheid als criterium, en wel voor leesindex A van Brouwer (1963; zie hoofdstuk 2). De gewenste leesbaarheid of moeilijkheidsgraad van de teksten was in hun visie afhankelijk van de leesvormen waarvoor de teksten gebruikt werden. Daarbij lieten ze zich leiden door Brouwer (1976) die op basis van een analyse van teksten die in het onderwijs werden gebruikt, kwam tot zogenaamde leesindexnormen voor de verschillende leesvormen. Het ter discussie stellen van de houdbaarheid van deze indeling valt buiten het bestek van deze paragraaf, en komt in het volgende hoofdstuk aan de orde. Wel geven we alvast ter illustratie de indeling van Brouwer weer in tabel 1.2.

Tabel 1.2: Normen leesindex A voor de verschillende leesvormen (overgenomen uit Brouwer, 1976, p. 501)

klas		Leesindex A				
aanvankelijk lezen	1	≥ 100				
voortgezet lezen		technisch lezen	voordracht lezen	belangstellend lezen	begrijpend lezen	studerend lezen
	2	90-99	90-99	100-109	90-99	90-99
	3	80-89	80-89	90- 99	80-89	80-89
	4	70-79	70-79	80- 89	70-79	70-79
	5	70-79	70-79	70- 79	60-69	60-69
	6	70-79	70-79	60- 69	50-59	50-59

N.B. hoe hoger de leesindex, hoe makkelijker de tekst

Speciaal voor het onderwijs in technisch lezen creëerde men negen AVI-niveaus, vier voor het aanvankelijk en vijf voor het voortgezet technisch lezen. Bij de samenstelling van de eerste vier AVI-niveaus hebben de doelstellingen voorop gestaan die Van den Berg en Te Lintelo (1977b) onderscheidden voor het aanvankelijk leesonderwijs, van doel 1 "de associatie van afbeelding met schriftbeeld en klankbeeld van concrete grondwoorden" tot doel 12 "het kunnen lezen van drie- en meerlettergrepige woorden" (p.57-59). De overige vijf AVI-niveaus baseerden ze op leesindex A. We zullen in hoofdstuk 2 uitge-

breid terugkomen op deze niveau-indeling. Bijlage 1 bevat de beschrijving van de negen AVI-niveaus.

De negen AVI-niveaus werden en worden enerzijds gebruikt om leesboeken en teksten in te delen naar moeilijkheidsgraad, waartoe de AVI-niveaulijsten voor het technisch lezen beschikbaar zijn (Assenberg, 1992; Assenberg & Schuurman-Hess, 1989). Anderzijds worden ook de prestaties op technisch lezen van de kinderen ingedeeld naar deze negen niveaus. Daarvoor zijn de AVI-toetskaarten ontwikkeld. Wanneer een leerling op een bepaald AVI-niveau is ingedeeld dan wordt daarmee aangegeven dat die leerling - in technisch opzicht - leesstof van een bepaalde moeilijkheidsgraad aankan. Op deze manier kan een leerkracht snel en eenvoudig te weten komen hoe goed de leerlingen technisch lezen en kan de leerkracht vervolgens, met behulp van de niveaulijsten, de leerlingen die stof laten lezen die afgestemd is op hun niveau van technisch lezen (zie ook Van Laarhoven, Peterink & Winnips, 1986).

Een veel gehoorde kritiek is dat het kiezen van leesstof op basis van AVI-niveaus voorbijgaat aan het belang van bijvoorbeeld de inhoud van leesboeken (Freeman-Smulders, 1995). Deze kritiek is onterecht. In het AVI-pakket is het AVI-niveau nooit als het enige criterium gezien voor de selectie van leesboeken. De inhoud, de herkenbaarheid en de opbouw van een boek, de afstemming op de gebruikte leesmethode en de interesse van het kind vond men bij die selectie zeker zo belangrijk (zie Van den Berg & Te Lintelo, 1977b, p. 26-27, p. 30 en p. 117). Naast de indeling in AVI-niveaus bevatte het pakket ook een lijst met inhoudelijke criteria (p. 37-38). De selectie van leesstof op grond van de AVI-niveaus, voor het technisch leesonderwijs, is wel het meest concreet uitgewerkt, terwijl het selecteren van leesstof voor de doelstellingen leesbegrip en leesbeleving in het pakket slechts globaal werd aangegeven. Daardoor is in de loop der tijd onterecht het idee ontstaan dat het AVI-niveau het enige criterium is voor de selectie van leesstof voor alle toepassingen in het leesonderwijs.

Samenvattend kunnen we vaststellen dat het AVI-pakket van oorsprong bedoeld was voor het totale leesonderwijs in de basisschool, zowel voor de leesteknik, het leesbegrip als de leesbeleving. Maar vooral de uitwerkingen voor het technisch lezen zijn in de praktijk ingevoerd en worden nog steeds gebruikt, waaronder de AVI-toetskaarten, de niveaulijsten voor het technisch lezen en de boekenlijsten voor moeilijk lezenden (zie Van den Berg, 1991). Daarbij zal zeker het feit dat voor het technisch leesonderwijs een en ander zeer concreet was uitgewerkt, een belangrijke rol hebben gespeeld.

1.1.2 Een beschrijving van de AVI-toetskaarten

Het doel van de AVI-toetskaarten is het niveau van technisch lezen van kinderen vast te stellen om zo, ten behoeve van het technisch leesonderwijs, leesstof te kunnen kiezen die aansluit bij hun niveau.

De AVI-toets bestaat uit negen kaarten, voor elk leesniveau één. Van de AVI-toets zijn een A- en een B-versie beschikbaar. Elke kaart bevat een tekst die een afgerond verhaal vormt (zie voor voorbeelden bijlage 2).

De afname van de toets vindt individueel plaats. De leerling krijgt een kaart voor zich met daarop een tekst behorend bij het niveau waarop hij of zij geacht wordt te lezen. De

illustratie achterop de kaarten voor de niveaus 1 tot en met 4 kan de leerkracht gebruiken om een gesprekje te houden met de leerling teneinde deze op zijn gemak te stellen. De leerling moet de tekst hardop voorlezen. De leerkracht noteert het aantal fouten dat de leerling maakt en de tijd (in seconden) die de leerling nodig heeft om de tekst te lezen. Een woord kan slechts één keer fout worden geteld. Het volgende dient de leerkracht *fout* te rekenen:

- een verkeerd woord lezen, zonder het alsnog te verbeteren;
- een woord spellend lezen, zonder het daarna alsnog in één keer te lezen;
- een woord overslaan;
- woorden omwisselen;
- woorden tussenvoegen;
- als het oplezen van een woord langer dan vijf seconden duurt, in welk geval het woord wordt voorgezegd.

Niet fout zijn:

- verkeerde klemtoon;
- versprekingen en haperingen, waarna het woord alsnog goed wordt gelezen;
- een woord wordt goed gelezen, nadat het eerst gespeld is;
- een fout woord wordt verbeterd door het goede woord;
- dialectuitspraak;
- ze i.p.v. zij, je i.p.v. jij, we i.p.v. wij, z'n i.p.v. zijn.

Met behulp van een normentabel (zie tabel 1.3) waarin zowel de leestijd als het aantal fouten is verwerkt, kan de leerkracht de leesprestatie, ofwel het AVI-niveau, van de leerling bepalen. Telkens moet aan *beide* criteria voldaan zijn. Bijvoorbeeld om een score *goed* op kaart 1a te halen, moet een leerling de tekst lezen in maximaal 1'50", waarbij hij of zij hooguit één fout mag maken.

Wanneer de prestatie van de leerling *voldoende* is, kan worden gestopt met de afname en wordt de leerling ingedeeld op het betreffende AVI-niveau. Bij een *onvoldoende* prestatie krijgt de leerling de kaart van een niveau lager voorgelegd, bij een *goede* leesprestatie de kaart van een niveau hoger. De leerkracht gaat op deze wijze verder totdat de prestatie *voldoende* is bereikt. In de regel komt de standaardprocedure erop neer dat de leerkracht de leerling bij een *goede* score telkens een kaart van één niveau hoger laat lezen totdat de score van de leerling op een kaart *voldoende* is. Het AVI-niveau is dan het niveau van de kaart waarop de score *voldoende* is behaald.

Tabel 1.3: Normentabel oude AVI-toetskaarten, versie A (uit Van Laarhoven, Peterink & Winnips, 1986)

kaart	GOED, kan hoger		VOLDOENDE		ONVOLDOENDE, lager	
	leestijd <i>maximaal:</i>	aantal fouten <i>maximaal:</i>	leestijd <i>maximaal:</i>	aantal fouten <i>maximaal:</i>	leestijd <i>meer dan:</i>	aantal fouten <i>meer dan:</i>
1A	1'50"	1	2'10"	3	2'10"	3
2A	1'30"	1	1'45"	4	1'45"	4
3A	1'25"	1	1'40"	3	1'40"	3
4A	1'15"	2	1'30"	5	1'30"	5
5A	1'05"	2	1'20"	4	1'20"	4
6A	1'25"	3	1'40"	7	1'40"	7
7A	1'15"	3	1'30"	6	1'30"	6
8A	1'35"	2	1'50"	6	1'50"	6
9A	1'15"	2	1'30"	4	1'30"	4

1.1.3 Vernieuwing van de AVI-toetskaarten

De AVI-toets wordt veel gebruikt door de Nederlandse basisscholen: 60% gebruikt de toetskaarten een of meerdere malen per jaar (Touwen, 1990). De gebruikers van AVI, basisscholen en schoolbegeleidingsdiensten, gaven in het onderzoek van Touwen aan dat de teksten en illustraties - ontwikkeld in de jaren zeventig - niet meer afgestemd waren op de belevingswereld van de kinderen. Een herziening werd door hen als wenselijk gezien.

Een tweede reden om de AVI-toets te herzien lag in het feit dat de criteria voor de opbouw in moeilijkheidsgraad voor de niveaus 1 tot en met 4 vrijwel uitsluitend gebaseerd waren op de oude versie van de methode Veilig Leren Lezen. Gezien het marktaandeel van deze methode destijds ($\pm 85\%$) is dit wel verklaarbaar en verdedigbaar, maar inmiddels zijn veel nieuwe methoden voor aanvankelijk lezen op de markt verschenen. De indeling van de eerste vier niveaus diende daarom opnieuw bekeken te worden.

Andere kritiek betrof het ontbreken van een verantwoording over de wijze waarop de kaarten en de normen tot stand waren gekomen, de representativiteit van de normen en het niet beschikbaar zijn van gegevens over de validiteit en de betrouwbaarheid van de toets (Dickhout-Rutten, 1983; Dickhout-Rutten & Mommers, 1984).

In de eerste publikaties over de AVI-toets (Van den Berg & Te Lintelo, 1977a; 1977b) is nauwelijks iets terug te vinden over de wijze waarop de teksten en de normen tot stand zijn gekomen. Met de term 'normen' worden de criteria voor de leestijd en de fouten bedoeld die bepalen bij welk AVI-niveau een leerling wordt ingedeeld op basis van zijn

prestaties op een toetskaart (vergelijk tabel 1.3). Uit de beschrijving van Dickhout-Rutten (1983) blijkt dat de oorspronkelijke tijd- en foutennormen voor de niveaus 1 tot en met 4 waren vastgesteld in een ijkingsonderzoek in 1976 onder 200 leerlingen afkomstig uit het bijzonder onderwijs in de regio Enschede. De normen voor de overige vijf niveaus had men daaruit afgeleid.

Uit het onderzoek van Dickhout-Rutten (1983) naar de betrouwbaarheid en validiteit van de AVI-toetskaarten (versie A) bleek dat de oorspronkelijke indelingscriteria niet voldeden. Zo vertoonde de frequentieverdeling van de proefpersonen (afkomstig uit de toenmalige klassen 1 tot en met 3) over de AVI-niveaus twee pieken. In alle onderzochte klassen lazen relatief weinig leerlingen op niveau 6 en 8 en relatief veel leerlingen op niveau 7. Daarnaast bleek dat bij opsplitsing van de AVI-norm in een tijds- en een foutennorm de zwaarte van die normen verschilde: bij de kaarten 2, 3, 4, 5, 8 en 9 woog de tijdsnorm zwaarder bij het bepalen van het niveau, terwijl bij de kaarten 1, 6 en 7 de foutennorm het zwaarste woog.

Dickhout-Rutten heeft, op basis van gegevens afkomstig van het project 'Preventie van Leesmoeilijkheden' (Van Dongen, Hulsmans, Voeten & Mommers, 1983), getracht de normering te herzien, althans voor de A-versie, waarbij ze de Eén-Minuut-Test (Brus & Voeten, 1973) als ijkingspunt heeft gekozen. Deze herziening is door het KPC overgenomen (Van Laarhoven, Peterink & Winnips, 1986).

De ontevredenheid over de normen van de AVI-toetskaarten blijkt uit het feit dat in de loop van de tijd verschillende normen voor de toegestane leestijd en het toegestane aantal fouten in omloop zijn gebracht. Touwen gaf aan dat 43% van de scholen de normeringen van de toetsontwikkelaar (het KPC) gebruikt (de oude of de door Dickhout-Rutten herziene A-normen). Meer dan de helft gebruikt dus andere normen (veelal van schoolbegeleidingsdiensten of van onbekende herkomst). Tevens worden verschillende scoringswijzen gehanteerd: wat de een 'fout' telt, rekent een ander 'half fout' of zelfs 'goed' (zie bijvoorbeeld Faber, 1989, 1990). Deze praktijk leidt ertoe dat leerlingen met exact dezelfde prestaties op een toetskaart (leestijd en fouten) ingedeeld worden in verschillende AVI-niveaus, afhankelijk van de gebruikte criteria.

Bovenstaande redenen hebben uiteindelijk geleid tot de herziening van de AVI-toetskaarten (Visser, Van Laarhoven & Ter Beek, 1994). In hoofdstuk 2 wordt deze herziening besproken. In de volgende paragrafen zullen we nagaan welke factoren een rol spelen in het proces van technisch lezen en de verwerving van de technische leesvaardigheid om zo te komen tot een formulering van eisen waaraan een toets voor technisch lezen dient te voldoen.

1.2 Technisch lezen

Hoe komt een lezer van een gedrukte tekst tot begrip? Thomassen, Noordman en Eling (1991) geven een kort overzicht van de psychologische processen die tijdens het lezen actief zijn: de visuele waarneming van letters, het begrijpen van woorden, het begrijpen van de structuur en de betekenis van de zin, en het begrijpen van zinnen en tekst. Daar-

bij gebruiken we onze kennis over orthografie, klanken, woorden, syntactische functie, betekenis, tekststructuur en het betreffende onderwerp.

We concentreren ons in deze paragraaf op de wijze waarop de lezer een relatie legt tussen het geschreven woord en de betekenis van dat woord in het geheugen, in het mentale lexicon. Dit wordt ook wel het decoderen genoemd, waarbij we decoderen opvatten in de brede betekenis van woordherkenning zoals door Goodman beschreven: het vertalen van een grafische input naar een betekeniscode (Samuels & Kamil, 1984). Goodman was de eerste die de term decoderen in deze betekenis gebruikte. Anderen gebruikten de term voor het proces waarin de grafische input wordt omgezet in fonemische input, waarvoor Goodman de term 'recoding' gebruikt.

In dit proefschrift hanteren we de in het Nederlandse onderwijs ingeburgerde term 'technisch lezen': het ontsleutelen van gedrukte of geschreven informatie (Brus & Voeten, 1973). De termen 'decoderen' en 'woordherkenning' gebruiken we als synoniemen van 'technisch lezen'. 'Technisch lezen' wordt in het onderwijs gebruikt ter onderscheiding van het 'begrijpend lezen': het kennen, afleiden en beoordelen van de betekenis van schriftelijke informatie (Aarnoutse, Mommers, Smits & Van Leeuwe, 1986). We sluiten ons aan bij Gough: "Word recognition is the foundation of the reading process" (1984, p.225). Maar tegelijkertijd zijn we ons er van bewust dat: "to emphasize the centrality of word recognition is not to deny that the ultimate purpose of reading is comprehension" (Stanovich, 1991, p.418).

Een veelgebruikte ingang om de discussie binnen de theorie en het onderzoek naar het leesproces weer te geven is die van de verschillende modellen voor de beschrijving van het leesproces: top-down, bottom-up en interactief. Voor deze ingang hebben we niet gekozen, alhoewel de modellen wel ter sprake zullen komen. In aansluiting op de reviews van Gough (1984) en Stanovich (1991) concentreren we ons op vier thema's die in de discussie rond woordherkenning belangrijk zijn:

- de perceptuele eenheid van verwerking;
- de rol van fonologische processen;
- de rol van de context;
- woordherkenning als een automatisch proces.

De eerste twee thema's betreffen vooral het proces van woordherkenning zelf, de andere twee thema's gaan tevens in op de relatie tussen woordherkenningsprocessen en hogere orde processen, zoals het begrijpen. Deze thema's houden direct verband met manieren om de technische leesvaardigheid te meten. Kunnen we de technische leesvaardigheid vaststellen door na te gaan of kinderen letters kunnen herkennen of moeten ze hele woorden kunnen lezen (perceptuele eenheid van verwerking)? Staat technisch lezen gelijk aan het kunnen toepassen van de letter-klank-koppeling, of is er meer aan de hand (de rol van fonologische processen en van de context)? Op welke manier stellen we vast dat het technisch lezen is geautomatiseerd?

Voor we overgaan naar het bespreken van de vier thema's willen we opmerken dat in deze paragraaf bij de beschrijvingen en verklaringen van het leesproces nogal eens uitgegaan wordt van het proces zoals zich dat afspeelt bij een ervaren lezer. Waar nodig zullen we aangeven op welke punten sprake is van verschillen tussen beginnende en

ervaren lezers. De ontwikkeling van de technische leesvaardigheid bij beginnende lezers zal in paragraaf 1.3 aan de orde komen.

1.2.1 De eenheid van verwerking bij woordherkenning

De eerste vraag is die naar de perceptuele eenheid van verwerking: herkennen we woorden op basis van letters, letterclusters, lettergrepen, woorden of nog grotere gehelen? Op grond van gezond verstand zouden we zeggen dat de letter de eenheid van verwerking is (Gough, 1984). Immers, in onze taal zijn de woorden samengesteld uit letters en veel woorden verschillen van elkaar op grond van slechts één letter.

In de *bottom-up* modellen van het leesproces wordt uitgegaan van een strikt lineaire opvatting over het proces van woordherkenning, waarbij het leesproces letter-voor-letter, woord-voor-woord verloopt (Seegers, Aarnoutse & Mommers, 1987). Met toenemende ervaring worden geen stappen overgeslagen, maar verlopen de diverse deelstappen steeds sneller. Gough beschrijft in zijn model van het leesproces (1972) de verwerking van een tekst vanaf het moment dat de ogen naar de tekst kijken tot het afleiden van betekenis. In het kort komt het model erop neer dat via oogfixaties de lezer een woord letter voor letter identificeert, van links naar rechts. Vervolgens wordt hiervan een fonemische code ofwel klankcode gemaakt. Via deze fonemische code wordt toegang verkregen tot het lexicon, en dus tot de betekenis van het woord. Ook de andere woorden in een zin worden serieel van links naar rechts, begrepen. Wel worden de afzonderlijke woorden van een zin opgeslagen in het werkgeheugen totdat ze in een grotere eenheid (zin) kunnen worden georganiseerd.

In de *top-down* modellen van het leesproces gaat men er daarentegen vanuit dat ervaren lezers grotere eenheden dan letters gebruiken voor woordherkenning. Goodman is een belangrijke exponent van deze benadering: "Reading is a psychological guessing game" (Goodman, 1976b, p. 498). Kenmerkend voor zijn model is dat de lezer zoveel mogelijk uitgaat van bestaande linguïstische (syntactisch en semantische) informatie en zo min mogelijk van de grafische informatie en de kennis over letter-klank-koppelingen.

In Goodmans optiek (1973) hangt de eenheid van woordherkenning af van het stadium van de leesontwikkeling waarin de lezer zich bevindt. Bij toenemende leeservaring vindt het lezen steeds meer van boven naar beneden plaats. Op het eerste niveau (beginnende lezer) wordt een woord letter-voor-letter gelezen, en komt de lezer via de klankcodering tot de interpretatie ervan. In het volgende stadium zal de lezer van grotere eenheden dan letters gebruik maken als grafische input, bijvoorbeeld letterclusters en woordvormen. In het laatste stadium worden nog grotere eenheden verwerkt (woordpatronen, woordgroepen, zinsdelen, semantische en syntactische informatie) en verdwijnt ook de invloed van de klankcode bij het lezen, waarover meer in paragraaf 1.2.2.

Deze gedachtengang vinden we ook terug bij Smith (1973; zie ook Nicholson, 1992 en Stanovich, 1986) die benadrukt dat een ervaren lezer niet alles wat op een pagina staat nodig heeft om de tekst te begrijpen.

Duidelijk is dat er nogal wat twijfel ontstond over de letter als perceptuele eenheid van verwerking. Al in 1886 sprak Cattell zijn twijfels uit over de rol van letters als perceptuele eenheid van woordherkenning (zie Gough, 1984 en Venezky, 1984). In zijn experimen-

ten vond hij namelijk dat woorden net zo snel gelezen konden worden als een afzonderlijke letter, en soms zelfs sneller. Deze bevinding ondergraaft echter niet direct de hypothese dat de letters de eenheid van verwerking zijn. Het resultaat kan verklaard worden door bijvoorbeeld de werking van het geheugen: men onthoudt dingen beter die aan elkaar gerelateerd zijn (zie Gough, 1984).

Een ander onderzoeksresultaat op grond waarvan getwijfeld werd aan de letter als eenheid van verwerking betreft het feit dat woorden herkend kunnen worden terwijl de afzonderlijke letters niet (volledig) herkend konden worden (Gough, 1984), bijvoorbeeld woorden waarvan alleen de onderste helft zichtbaar is. Op grond van het feit dat een goede lezer een verminkte tekst *kan* lezen, mogen we echter niet concluderen dat het *normale* leesproces op die manier plaatsvindt (Nicholson, 1992).

Reicher (1969) vond dat herkenning van een letter die binnen een woord wordt aangeboden sneller verloopt, dan wanneer die letter afzonderlijk wordt aangeboden; het zogenaamde woordsuperioriteitseffect. Echter, ook deze bevinding bleek uiteindelijk geen bewijs tegen de letter als eenheid van verwerking. Het effect bleek mede een gevolg te zijn van de gebruikte onderzoeksmethode. En nog veel belangrijker was het dat het effect ook aanwezig was bij het aanbieden van letters in pseudowoorden (zie Gough, 1984).

In hedendaagse modellen voor woordherkenning gaat men er vrijwel unaniem van uit dat tijdens de woordherkenning representaties worden aangesproken voor eenheden die kleiner zijn dan het hele woord. De aard van die eenheden staat echter ter discussie; mogelijke kandidaten zijn letters, syllaben, morfemen, bigrammen en trigrammen (dit zijn combinaties van twee respectievelijk drie vaak naast elkaar voorkomende letters, bijvoorbeeld 'st' of 'str') (Grainger & Dijkstra, 1991; Perfetti, 1984). Het bewijs dat grotere orthografische eenheden dan grafemen worden verwerkt tijdens het lezen is echter gering; wel blijkt dat het orthografische patroon van een woord een rol speelt tijdens woordherkenning (Reitsma, 1990; Theloozen & Van Bon, 1993).

Voor woordherkenning via letterperceptie pleit de bevinding in onderzoek naar oogbewegingen dat de ervaren lezer vrijwel ieder woord fixeert wanneer hij leest, en dat het oog elke individuele letter lijkt te verwerken (Nicholson, 1992; Stanovich, 1986). In twee veel besproken modellen voor woordherkenning, het interactieve activatiemodel van McClelland en Rumelhart (zie Grainger & Dijkstra, 1991, en Perfetti, 1984), en het activatie-verificatiemodel van Paap, Newsome, McDonald en Schvaneveldt (zie Grainger & Dijkstra, 1991), wordt de afzonderlijke letter expliciet gebruikt als eenheid die een rol speelt tussen de kenmerkanalyse en de toegang tot het lexicon. Beide modellen kunnen het eerder genoemde woordsuperioriteitseffect verklaren door aan te nemen dat meerdere letters tegelijkertijd geïdentificeerd worden (in plaats van letter-voor-letter) en dat woorden herkend kunnen worden voordat de identificatie van alle letters is afgelopen.

Vanuit onze optiek, gericht op het meten van technisch lezen, is de les die we leren uit deze paragraaf dat letteridentificatie belangrijk is voor het technisch lezen (en naar later zal blijken, zeker voor het leren lezen), maar dat de onderzoeksresultaten er ook op wijzen dat bij woordherkenning grotere eenheden een rol kunnen spelen, met name orthografische patronen.

1.2.2 De rol van fonologische processen

De rol van fonologische processen is in de vorige paragraaf reeds even ter sprake gekomen bij het bespreken van de top-down en bottom-up opvattingen van het leesproces. In vrijwel alle huidige theorieën over het leesproces gaat men er vanuit dat fonologische codes een rol spelen bij het *begrijpen van zinnen*: de klankcode vergemakkelijkt het onthouden van betekenissen in het werkgeheugen (Perfetti, 1985; Stanovich, 1991). De vraag die we ons hier stellen is: verloopt *woordherkenning* via een fonologische code? Met andere woorden: spelen fonologische processen prelexicaal een rol?

De fonologische hercoderingshypothese van Rubinstein, Lewis en Rubinstein (zie Gough, 1984, en Stanovich, 1991) houdt in dat woordherkenning plaatsvindt via een fonologische vorm. Woordherkenning verloopt dan in twee fasen: eerst worden de letters omgezet in fonemen, de klankcode, en vervolgens wordt in het mentale lexicon gezocht naar de betekeniscode die bij de betreffende klankcode hoort. In deze opvatting speelt dus een of andere klankcode prelexicaal een cruciale rol bij herkenning van een woord. Overigens moet die klankcode niet als spraak of subvocalisatie (in jezelf praten) worden gezien, maar als een abstracte reeks fonemen (Perfetti, 1985).

Deze hypothese past goed in de bottom-up benadering, volgens welke de ervaren lezer ieder woord letter voor letter omzet in klanken en via die klankcode toegang krijgt tot het mentale lexicon. In de vorige paragraaf zagen we dat in het top-down model van Goodman aan de klankcode wel een rol wordt toegekend bij woordherkenning van de beginnende lezer, maar dat uiteindelijk de ervaren lezer direct een woord herkent, zonder tussenkomst van een klankcode.

Samenvattend kunnen we vaststellen dat volgens de strikte bottom-up benadering woordherkenning altijd via een fonologische code plaatsvindt, terwijl volgens de top-down benadering de ervaren lezer vanuit de visuele code direct toegang heeft tot het lexicon.

In de eerste stadia van verwerving van de leesvaardigheid blijken fonologische processen een belangrijke rol te spelen (zie Stanovich, 1991). We zullen daarop terugkomen in paragraaf 1.3. Onderzoek wees echter uit dat de sterke versie van de fonologische hercoderingshypothese van Rubinstein et al. (woordherkenning verloopt altijd via de klankcode) niet gehandhaafd kan worden voor het leesproces van een ervaren lezer (zie Gough, 1984; Stanovich, 1991). Men ontwikkelde vervolgens de zogenaamde 'dual route'-modellen, waarin men ervan uitgaat dat er twee alternatieve routes bestaan naar het lexicon, namelijk een directe route (visueel-orthografisch) en een indirecte via de fonologische codering. Welke route gevolgd wordt tijdens het lezen is afhankelijk van bijvoorbeeld de frequentie en de regelmatige spelling van een woord. Een voorbeeld van deze opvatting is Colthearts 'horse race model' (zie Gough, 1984).

In deze benadering blijft zowel de top-down benadering (directe toegang) als de bottom-up benadering (via fonologische code) gehandhaafd. Dit kan worden verklaard door het feit dat beide manieren van woordherkenning ondersteund worden door onderzoeksresultaten (Van den Broek, 1993). Wat pleit voor het bestaan van een indirecte route zijn de bevinding dat pseudowoorden moeiteloos gelezen kunnen worden en het bestaan van een regelmatigheidseffect (regelmatig gespelde woorden worden doorgaans sneller gelezen). Voor het bestaan van een directe route pleit het lezen van onregelmatig

gespelde woorden (waarvoor de grafeem-foneem correspondentieregels tekortschieten), het woordsuperioriteitseffect, en het frequentie-effect (hoogfrequente woorden worden sneller gelezen dan laagfrequente).

Onderzoek van Seidenberg en anderen (zie Stanovich, 1991) toonde aan dat het eerder genoemde regelmatigheidseffect interacteert met het frequentie-effect. Het regelmatigheidseffect bleek niet of nauwelijks te bestaan voor hoogfrequente woorden: onregelmatige en regelmatige hoogfrequente woorden werden even snel gelezen. Het regelmatigheidseffect nam toe als de woordfrequentie afnam. Seidenberg leidde hieruit af dat in ieder geval voor laagfrequente woorden fonologische codering plaatsvindt voordat lexicaal toegang wordt verkregen.

Experimenten van Ghlusko gaven echter aanleiding om deze conclusie in twijfel te trekken (zie Perfetti, 1985 en Stanovich, 1991). Deze experimenten lieten het zogenaamde buurwoordfrequentie-effect zien: nabijgelegen lexicale toegangen beïnvloeden de snelheid van benoemen van woorden. Bijvoorbeeld het benoemen van regelmatig gespeelde woorden als het Engelse woord 'gave' die onregelmatige burens hebben (in dit geval 'have') kostte meer tijd dan het uitspreken van regelmatige woorden zonder onregelmatige burens (bijvoorbeeld 'coat'). Deze bevindingen leidden tot de ontwikkeling van zogenaamde lexicale-analogie-modellen, zoals bijvoorbeeld Ghlusko's 'activation-synthesis model'. Daarin veronderstelt hij geen afzonderlijk fonologisch codeermechanisme, maar gaat hij ervan uit dat de fonologische codering tot stand komt door het gebruiken van eerder opgeslagen lexicale toegangen. Herkenning van een woord vindt primair plaats via de visueel-orthografische code. De fonologische informatie wordt geactiveerd via de klankeigenschappen die met de orthografische structuur verbonden zijn. De fonologische code van een nieuw of laagfrequent woord wordt op automatische wijze geconstrueerd via analogie met de fonologische codes van orthografische burens (zie ook Seegers, 1985).

Ghlusko's model alsmede het eerder genoemde model van McClelland en Rumelhart zijn de voorlopers van de zogenaamde 'distributed processing' modellen, waarvan de modellen van Perfetti (1985; 'automatic speech activation'), en Seidenberg (het 'time course' model; zie Stanovich, 1991, 1992) voorbeelden zijn. Volgens deze modellen activeren letters, fonemen en woordcodes elkaar simultaan binnen de woordherkenningsmodule. Fonologische activatie is dan niet een eerste stap naar de toegang tot het lexicon, maar is deel van die toegang, waarbij het fonologische proces soms al voltooid is voordat de toegang volledig is en soms niet (Perfetti, 1985). De vraag naar activatie van de fonologische code voor of na de toegang tot het lexicon (pre- of postlexicaal) is in deze modellen vervangen door de vraag hoeveel fonologische mediatie plaatsvindt.

Beide modellen geven echter slechts verbale verklaringen van woordherkenning. Recentere modellen, de zogenaamde 'connectionistic distributed processing' modellen, geven meer expliciete rekenkundige verklaringen daarvoor via computersimulaties (bijvoorbeeld de modellen van Seidenberg & McClelland en Van Orden; voor meer informatie zie Stanovich, 1991 en 1992, en Van den Broek, 1993). In deze modellen gaat men niet uit van het bestaan van een afzonderlijke fonologische route of een afzonderlijk lexicon en veronderstelt men geen representaties op woordniveau. Een woord wordt op gedistribueerde

wijze voorgesteld in een netwerk van orthografische en fonologische kenmerken. Telkens als men een woord leest, wordt een specifiek patroon geactiveerd in datzelfde netwerk. De kennis van woorden is verwerkt in een set gewichten op verbindingen tussen verwerkingseenheden voor letters, fonemen en de correlaties daartussen. Volgens het leerprincipe van het covariante leren worden verschillen op het ene vlak (bijvoorbeeld spelling) die samenhangen met verschillen op een ander vlak (bijvoorbeeld fonologie) met elkaar in verband gebracht en op een of andere wijze opgeslagen. Een voorbeeld kan dit wellicht verhelderen. Als we bijvoorbeeld herhaaldelijk het woord 'in' lezen, dan leren we de relatie tussen de letter 'i' en de klank /i/, tussen 'n' en /n/, tussen de letters 'i' en 'n', tussen de klanken /i/ en /n/ en ook tussen 'i' en /n/ en tussen /i/ en 'n'. Als we daarna herhaaldelijk het woord 'ik' lezen, wordt de relatie tussen 'i' en /i/ nog verder versterkt; het gewicht op deze verbinding wordt hoger.

Zowel in deze 'connectionistische distributie-modellen' als in het model van Perfetti gaat men ervan uit dat snel toegang wordt verkregen tot fonologische informatie als een automatisch gevolg van orthografische activatie. Recent onderzoek (zie Stanovich, 1992) ondersteunt dit door de bevinding dat fonologische activatie vrijwel onmiddellijk optreedt tijdens woordherkenning.

Deze recente modellen weerspiegelen een andere kijk op de wijze waarop fonologische processen een rol spelen bij woordherkenning. Het idee is dat fonologische informatie snel opgeroepen wordt door orthografische activatie en een automatisch gevolg is van die activatie. In deze optiek is het onderscheid tussen prelexicaal en postlexicaal optreden van fonologische processen verdwenen.

Deze opvatting is echter nog niet algemeen geaccepteerd door onderzoekers; de dual-route modellen hebben ook nog steeds veel aanhang. Stanovich (1991, 1992) geeft aan dat één belangrijke bron van informatie ertoe leidt dat onderzoekers nog de dual-route modellen aanhangen, namelijk bevindingen ten aanzien van *verworven* dyslexie. Een deel van de mensen met verworven dyslexie kan alleen eerder geleerde woorden lezen, maar geen pseudowoorden; een ander deel daarentegen kan wel pseudowoorden en regelmatige woorden lezen, maar heeft problemen met onregelmatige woorden en homofonen. Dit zou kunnen wijzen op selectieve schade aan de indirecte route via grafeem-foneem regels voor de eerstgenoemde groep, respectievelijk aan de directe visuele route voor de tweede groep.

Waarover wel vrijwel iedereen het eens is, is de rol van de fonologische code bij het *leren* decoderen, waarover meer in paragraaf 1.3.

Voor het meten van technische leesvaardigheid kunnen we uit het bovenstaande afleiden dat kennis van de letter-klank-koppeling belangrijk is. Maar tegelijkertijd moeten we ons realiseren dat technisch lezen meer is dan alleen het koppelen van letters aan klanken. Het gebruik maken van fonologische processen lijkt slechts één manier te zijn om woorden te herkennen. Een andere belangrijke manier om woorden te identificeren is de visueel-orthografische strategie, waarbij het geschreven woord direct wordt vergeleken met representaties in het geheugen.

Ook vanuit een ander perspectief is het goed verdedigbaar om het meten van technisch lezen niet te beperken tot het meten van kennis van de letter-klank-koppeling. Immers

het Nederlandse schriftsysteem is niet alleen op het fonologische principe gebaseerd. Het morfologische principe en het etymologische principe zijn eveneens belangrijke pijlers van onze orthografie. Toepassing van deze twee principes leidt nogal eens tot schending van het fonologische principe (Assink, 1989). Voor het identificeren van woorden is daarom ook orthografische kennis noodzakelijk (Reitsma, 1983, 1990).

1.2.3 De rol van de context

Uiteraard speelt de context een belangrijke rol bij het begrijpen van een tekst, maar speelt de context ook pre-lexicaal een rol, bij het herkennen van woorden?

Lange tijd is de top-down opvatting ten aanzien van de rol van de context bij woordherkenning dominant geweest: verwachtingen en contextuele informatie zouden ook aan woordherkenning een grote bijdrage leveren. In de top-down modellen gaat men er vanuit dat de ervaren lezer voor woordherkenning sterk afhankelijk is van de context (zie Seegers et al, 1987). De ervaren lezer gebruikt de context om verwachtingen te formuleren ten aanzien van binnenkomende informatie. Als die informatie klopt met de verwachtingen vindt een beperkte analyse plaats, en richt de lezer zich op een volgend stuk informatie. "As a child develops reading skill and speed, he uses increasingly fewer graphic cues" (Goodman, 1976b, p.504). "But if in fact you are not making errors when you read, you are probably not reading efficiently, you are processing more visual information than you need" (Smith, 1979, p.33). Bevestiging van deze opvatting dacht men te vinden in onderzoek dat erop wees dat lezers leesfouten maken die qua grammatica en betekenis juist zijn, en dat lezers spelfouten niet opmerken, met name als woorden goed in de context passen (zie Seegers et al). Ook vond Goodman (1985) dat als leerlingen een woord in de context lezen, de leesnauwkeurigheid verbeterde met 60-80% ten opzichte van het lezen van een woordenlijst.

Een veelheid van onderzoek wijst er echter op dat bij de ervaren lezer de context pre-lexicaal, dus vóór woordherkenning, geen rol speelt.

Zo kunnen lezers wel globaal voorspellen welke informatie er zal komen, maar deze voorspellingen zijn niet voldoende specifiek om als basis te dienen voor een toetsend leesproces. Woordherkenning verloopt zo snel dat het de moeite niet loont om verwachtingen af te leiden (zie Samuels & Kamil, 1984). Onderzoek van Gough (1983) toonde aan dat één op de vier woorden goed voorspeld kan worden; voor inhoudswoorden geldt zelfs een verhouding van 1 op 10.

Uit onderzoeken naar oogbewegingen en oogfixaties bleek dat tijdens het lezen de overgrote meerderheid van de inhoudswoorden gefixeerd wordt (voor een overzicht zie Nicholson, 1992; Perfetti, 1984, 1985; Stanovich, 1980, 1982a, 1982b, 1986, 1991, 1992). Korte functiewoorden en hoog-voorspelbare woorden worden weleens een keer overgeslagen, maar meestal toch gefixeerd. Het aantal fixatiepunten is zeer hoog, hetgeen niet wijst op het overslaan van woorden, en ook lijken alle letters van een woord te worden gescand gedurende een fixatie. Deze bevindingen zijn in tegenspraak met het idee dat verwachtingen en voorspellingen het primaat hebben, en visuele informatie slechts ter bevestiging wordt gebruikt.

Replicatie van het eerder genoemde onderzoek van Goodman toonde aan dat het effect dat een woord beter in de context wordt gelezen veeleer het gevolg was van oefening.

De leerlingen lezen namelijk eerst de woordenlijst en vervolgens het verhaal waarin dezelfde woorden voorkwamen.

Onderzoek naar leesfouten en reactietijden bij hardop lezen toont aan dat zowel onervaren/zwakke lezers als ervaren/goede lezers gebruikmaken van de context, en de eersten zelfs iets meer (Stanovich, 1980, 1992). Deze bevinding dat onervaren/zwakke lezers zelfs iets meer gebruik maken van de context is eveneens in tegenspraak met de top-down opvatting, waarin men immers veronderstelt dat de ervaren lezers steeds sterker afhankelijk worden van de context.

Stanovich (1980) heeft de kennis over het gebruik van de context door ervaren en beginnende lezers verwerkt in zijn 'interactive-compensatory model'. Het sleutelbegrip daarin is dat een proces van het ene niveau kan compenseren voor deficiënties op het andere niveau (zie ook het 'verbal efficiency' model van Perfetti (1984, 1985)). De zwakke lezer, met een inaccurate en langzame woordherkenning maar met kennis over het onderwerp kan dit compenseren via top-down processing. Voor een ervaren lezer met weinig kennis over een onderwerp is het makkelijker bottom-up te werken. Dit model verklaart waarom juist zwakke lezers onder bepaalde omstandigheden gevoeliger zijn voor contextuele beperkingen, namelijk als de visuele, orthografische en/of lexicale kennisbronnen zwak zijn vergeleken met de syntactische en semantische kennis. Het niveau van de decodeervaardigheid bepaalt de mate waarin de lezer steunt op de context bij woordherkenning. Contextuele facilitatie kan echter alleen plaatsvinden als de context begrepen wordt (Perfetti, 1985). Een zwakke lezer zal juist door de zwakke decodeervaardigheid over minder contextuele informatie (afkomstig uit de tekst) beschikken om eventueel te kunnen compenseren voor die zwakkere decodeervaardigheid (Stanovich, 1986; Stanovich, Cunningham & Feeman, 1984). De ironie ervan ligt in het feit dat de betere lezers over betere contextuele informatie beschikken dan de zwakke lezers, maar die kennis niet of nauwelijks (hoeven te) gebruiken voor woordherkenning.

Het is overigens niet zo dat Stanovich in zijn model veronderstelt dat de woordherkenning van goede lezers niet vergemakkelijkt wordt door de voorafgaande context. In aansluiting op de twee-proces-theorie van Posner en Snyder (zie Stanovich, 1980) gaat hij ervan uit dat er twee soorten compenserende processen zijn waardoor de context een effect kan hebben op woordherkenning: een bewust aandacht vragend mechanisme en automatisch spreidende activatie. Simpson en Lorsch (1987) doelen op hetzelfde als ze onderscheid maken tussen bewuste en automatische contexteffecten. Het eerste context-effect is het gevolg van een langzaam proces waarin verwachtingen worden opgesteld en de aandacht bewust gericht wordt op bepaalde gebieden in het geheugen. Dit proces vraagt aandachtscapaciteit en werkt facilitair op de woordherkenning. Tegelijkertijd werkt het proces belemmerend op het ophalen van informatie uit een onverwachte plek in het geheugen, omdat in dat geval de aandacht naar een andere plaats in het geheugen verplaatst moet worden. Dit contexteffect is zichtbaar bij beginnende en zwakke lezers (Seegers, 1985; Stanovich, 1980, 1991).

Het tweede contexteffect treedt op als gevolg van de zogenaamde 'automatic spreading activation': door de herkenning van een woord treedt 'automatische' activatie op van de representatie in het geheugen en worden ermee verbonden representaties enigszins geac-

tiveerd. Dit contexteffect vindt dus plaats in het semantische geheugen, werkt snel en facilitair, maar niet belemmerend op het ophalen van informatie vanuit een onverwachte plek van het geheugen. Dit effect is terug te vinden bij volwassen lezers. Het idee van automatische spreidende activatie vertoont overeenkomst met het idee van Van Orden dat eventuele contexteffecten het gevolg zijn van versterkte connecties tussen samen variërende codes (zie Van den Broek, 1993). Veel onderzoek laat bovendien zien dat decoderen automatisch leidt tot semantische activatie, betekenisverlening, zelfs bij zwakke lezers, gegeven uiteraard dat de betekenis van het woord goed is opgeslagen in het geheugen (zie Stanovich, 1991).

De meeste recente bevindingen wijzen erop dat bij een ervaren lezer de context wel een belangrijke rol speelt bij het begrijpen van een tekst, maar pre-lexicaal, dat wil zeggen vóór woordherkenning, nauwelijks of geen rol speelt. De top-down opvatting dat ervaren lezers zo min mogelijk gebruik maken van visuele informatie en zoveel mogelijk van contextuele informatie is voor woordherkenning onhoudbaar gebleken. Het is niet zo dat goede lezers minder gebruik maken van visuele informatie, maar dat ze minder aandachtscapaciteit besteden aan het verwerken ervan. De reden dat ze er minder capaciteit aan besteden is niet omdat ze meer uitgaan van de context, maar omdat hun stimulus-verwerkingsmechanismen zo krachtig zijn.

Wat betekenen deze bevindingen nu als we technisch lezen willen meten? Kunnen we volstaan met het meten van woordherkenning door middel van het laten lezen van losse woorden? Of dienen we de woorden in een zinvolle context aan te bieden? Gezien de bevindingen dat de context bij een ervaren lezer nauwelijks of geen rol speelt bij woordherkenning zouden we bij het meten van technisch lezen kunnen volstaan met het aanbieden van losse woorden. In zo'n geval kan de beginnende lezer geen gebruik maken van contextuele informatie om de zwakke woordherkenning te compenseren, zodat we ook bij hen meten hoe efficiënt en snel hun woordherkenning verloopt. Wat echter pleit voor het aanbieden van samenhangende teksten is dat dit veel meer lijkt op de werkelijke leessituatie. De samenhangende tekst geeft degenen met een minder goed ontwikkelde woordherkenning dan wel de mogelijkheid te compenseren via het gebruiken van de context, maar gezien de vertragende werking van dit contexteffect zal dit niet snel leiden tot een verkeerde inschatting van de mate van beheersing van het technisch lezen.

1.2.4 Woordherkenning als een automatisch proces

In de vorige paragraaf zijn we tot de conclusie gekomen dat woordherkenning bij de ervaren lezer niet of nauwelijks beïnvloed wordt door hogere orde processen. Dit wijst erop dat woordherkenning een automatisch proces is. Het concept van geautomatiseerd-zijn is reeds in 1908 door Huey aan de orde gesteld (zie Stanovich, 1991), en is pas veel later weer opgepakt door LaBerge en Samuels (1974). Centraal in het model van LaBerge en Samuels staat het begrip 'aandacht'. In hun model veronderstellen ze dat aandacht vereist is om betekenis af te leiden uit een tekst en dat de totale aandachtscapaciteit beperkt is. Zowel het decoderen als het begrijpen vereist aandacht. Hoeveel aandacht vereist is voor het begrijpen van een tekst is afhankelijk van factoren als decodeervaar-

digheid, bekendheid met woorden in de tekst, het onderwerp, de dichtheid van ideeën in een tekst, enzovoorts. Voor een complexe vaardigheid als lezen is het noodzakelijk veel deelprocessen in een korte periode uit te voeren. Als ieder deelproces evenveel aandacht zou vragen, zou het uitvoeren van een complexe vaardigheid onmogelijk zijn omdat dan de aandachtscapaciteit overvraagd wordt. Vooral de prelexicale processen van woordherkenning (kenmerkanalyse, orthografische segmentatie en fonologische codering) komen in hun visie in aanmerking voor automatisering.

In de opvatting van LaBerge en Samuels zou woordherkenning automatisch verlopen als het proces geen aandachtscapaciteit vraagt. Om dit vast te stellen, maakten ze gebruik van een kenmerk dat samenhangt met capaciteitsvrije verwerking: automatische uitvoering of automatisme. Dit is de neiging van een geautomatiseerd proces om te werken ongeacht de taak waarop de bewuste aandacht van een subject gericht is. "Our criterion for deciding when a skill or subskill is automatic is that it can complete its processing while attention is directed elsewhere" (p.295). In hun optiek staat het ontbreken van bewuste aandacht gelijk met capaciteitsvrije verwerking.

Uit diverse onderzoeken bleek echter dat het automatisch/onbewust herkennen van woorden en capaciteitsvrije verwerking niet aan elkaar gelijk zijn, en dat deze ook niet hetzelfde zijn als snelle verwerking. Zo vraagt woordherkenning, ook al is het een automatisch proces, nog wel enige aandachtscapaciteit. Ook bleek dat laagfrequente woorden meer aandacht vragen dan hoog-frequente, terwijl ze beide 'automatisch' worden herkend. In andere onderzoeken werd aangetoond dat automatische/onbewuste uitvoering van woordherkenningsprocessen zich al vrij snel ontwikkelt, maar dat de snelheid en efficiëntie ervan, in termen van afnemend capaciteitsgebruik, zich verder ontwikkelen, ook nadat herkenning automatisch is geworden (Horn & Manis, 1987; Seegers, 1985; Stanovich, 1991).

Voor een goed begrip van woordherkenning als een automatisch proces is het dus belangrijk om capaciteitsgebruik te scheiden van snelheid en automatisme van woordherkenning.

Dit wordt bevestigd door onderzoek naar de ontwikkeling van woordherkenning. Ehri en Wilce (1979, 1983) vonden drie fasen in de ontwikkeling van woordherkenning, namelijk 1) de *nauwkeurigheid*sfase, waarin kinderen veel aandacht besteden aan het verwerken van letters en aan de letter-klank-koppeling, 2) de *automatische* fase, waarin ze woorden herkennen zonder aandacht voor de afzonderlijke letters en letter-klank-relaties, en 3) de *snelheid*sfase, waarin de snelheid van herkenning steeds verder toeneemt. En juist dat laatste, de steeds grotere snelheid waarmee woorden herkend worden, blijkt belangrijk te zijn voor het ontwikkelen van leesbegrip.

De nadere uitwerking van het concept 'automatisch' heeft ertoe geleid dat theorieën van lezen gebaseerd op capaciteitsgebruik minder nadruk kregen in onderzoek. Een verschuiving heeft plaatsgevonden van cognitief capaciteitsgebruik naar kennisrepresentaties (Stanovich, 1991). In de recentere opvattingen wordt een beginnende lezer niet zozeer beperkt door de beschikbare verwerkingscapaciteit, maar veel meer door gebrek aan beschikbare informatie (Logan, 1988a, 1988b; Näslund & Samuels, 1992). Door veel erva-

ring wordt een domein-specifiek kennisbestand opgebouwd, waardoor het steeds meer mogelijk wordt om te decoderen op grond van in het geheugen opgeslagen oplossingen. Het concept modulariteit van Fodor (zie Stanovich, 1991) heeft bij bovengenoemde verschuiving een grote invloed gehad. Kenmerkend voor een modulair proces is 'information encapsulation': eerder opgeslagen kennis beïnvloedt het betreffende proces niet. Snelle verwerking, domein-specificiteit en automatische uitvoering zijn eigenschappen die vaak samengaan met informatie-inkapseling. Een modulair proces is feitelijk een autonoom functionerend proces. Dit betekent ook dat zo'n verwerkingsmodule niet gecontroleerd wordt door hogere orde processen of aangevuld wordt door informatie van kennisstructuren buiten de module. Doordat modulaire processen geïsoleerd werken, kunnen ze informatie coderen zonder gestoord te worden en kunnen ze dus ook snel werken. Juist het proces van woordherkenning lijkt in aanmerking te komen voor 'information encapsulation', mede gezien de bevindingen in de vorige paragraaf ten aanzien van contexteffecten. Stanovich concludeert dan ook: "In short, an advantage accrues to encapsulation when the specificity and efficiency of stimulus-analyzing mechanisms is great relative to the diagnosticity of the background information that might potentially be recruited to aid recognition." (Stanovich, 1991, p.430). Immers, woordherkenning wordt steeds minder beïnvloed door de context als de leesvaardigheid toeneemt; ofwel, woordherkenning wordt steeds meer een autonoom functionerende verwerkingsmodule. Als deze module in gang wordt gezet door de binnenkomende grafische informatie, wordt deze informatie verwerkt door informatie die in de module zelf opgesloten zit, en wordt de verwerking niet beïnvloed door informatie vanuit andere, hogere orde processen. Wel vindt uitwisseling met hogere orde processen plaats na afloop van de verwerking, wanneer het resultaat wordt doorgegeven.

Mogelijke kleine contexteffecten bij de ervaren lezer zijn het gevolg van automatische spreidende activatie in het semantische geheugen. Bij de beginnende lezer zal de woordherkenning nog geen autonoom functionerende module zijn, waardoor, juist door het trage verloop van de woordherkenning, achtergrondkennis en verwachtingen kunnen doordringen in de woordherkenningsmodule.

Vanuit het perspectief van het meten van technische leesvaardigheid kunnen we uit de huidige inzichten over woordherkenning als automatisch proces duidelijke aanwijzingen halen. Om vast te stellen in welke mate het technisch lezen geautomatiseerd is, kunnen we niet volstaan met vast te stellen of een leerling nauwkeurig (zonder fouten) of automatisch (zonder bewuste aandacht) woorden kan herkennen. Deze twee aspecten zijn de eerste stappen op weg naar woordherkenning als automatisch proces. De snelheid van woordherkenning is het meest belangrijke te meten aspect. In de volgende paragrafen over de verwerving van de technische leesvaardigheid zal blijken dat juist de snelheid het langst voor verbetering vatbaar is en het meest belangrijk is voor het ontwikkelen van leesbegrip.

1.3 Hoe verwerven kinderen de technische leesvaardigheid?

In de vorige paragraaf hebben we vier belangrijke thema's omtrent woordherkenning elk vanuit verschillende opvattingen besproken. De besproken modellen hebben veelal betrekking op de technische leesvaardigheid van een ervaren lezer. We hebben hier en daar al aangegeven waar het proces bij een beginnende lezer anders zal zijn. Zo is al ter sprake gekomen dat bij ervaren lezers waarschijnlijk grotere eenheden dan de letter een rol spelen bij woordherkenning, maar dat letteridentificatie bij beginnende lezers wel degelijk belangrijk is (paragraaf 1.2.1). Ook zagen we dat de invloed van fonologische processen bij woordherkenning kan verschillen bij beginnende en ervaren lezers (paragraaf 1.2.2). Beginnende lezers hebben nog niet voldoende orthografische kennis opgebouwd om woorden te verwerken via een niet-gemedieerde, niet-fonologische route. Daarnaast lijken beginnende lezers voor woordherkenning veel meer afhankelijk te zijn van de context dan ervaren lezers (paragraaf 1.2.3). De laatsten lezen voornamelijk op basis van de grafische informatie. Tot slot zagen we dat vooral de snelheid van woordherkenning de ervaren lezer kenmerkt, terwijl een beginnende lezer deze snelheid, naast nauwkeurigheid en automatisme, nog moet verwerven (paragraaf 1.2.4).

In deze paragraaf besteden we aandacht aan de wijze waarop het technisch lezen plaatsvindt bij een beginnende lezer en hoe deze vaardigheid zich ontwikkelt. Dit doen we om te weten te komen waarmee we rekening moeten houden bij het meten van technisch lezen.

In paragraaf 1.3.1 bespreken we enkele modellen voor de verwerving van de leesvaardigheid, met name de technische leesvaardigheid. Daarbij komen de fasen aan bod die men onderscheidt in de ontwikkeling van de technische leesvaardigheid. In paragraaf 1.3.2 schenken we aandacht aan de wijze waarop een kind de technische leesvaardigheid verwerft in het onderwijs, waarbij ook de rol van de AVI-toets aan de orde komt.

1.3.1 Modellen voor de verwerving van de technische leesvaardigheid

De bestaande modellen voor het *verwerven* van de leesvaardigheid zijn gebaseerd op één van de twee volgende paradigma's (Juel, 1991).

In het *eerste paradigma* gaat men er vanuit dat het leesproces bij beginnende en ervaren lezers hetzelfde is. Beide typen lezers zijn gericht op het zoeken naar betekenis, waarbij kennis van de taal en de wereld (syntactische en semantische informatie) belangrijker is dan grafische informatie. De kwantitatieve groei in kennis over de taal en de wereld zijn de belangrijkste factoren die een beginnende en ervaren lezer van elkaar onderscheiden. *De lezer is het meest succesvol als een minimum aan orthografische kennis wordt gebruikt.* Het is duidelijk dat dit eerste paradigma aansluit bij de eerder besproken top-down modellen van het leesproces (Goodman, 1973, 1976a, 1976b; Smith, 1973): het leesproces als een psycholinguïstisch 'gokspel' waarin de lezer drie typen informatie (grafische, syntactische en semantische) tegelijkertijd verwerkt en coördineert. De resulterende modellen voor het verwerven van leesvaardigheid zijn zogenaamde 'non-stage'-modellen. In deze modellen wordt de leesontwikkeling gezien als een natuurlijk proces, dat parallel loopt aan de taalontwikkeling. Goodman en Smith zijn dan ook geen voorstanders van het leren decoderen waarbij de nadruk wordt gelegd op het leren van de grafeem-fo-

neem-koppeling. Dit zou namelijk leiden tot het fenomeen 'word-calling'. Het gebruik van deze term gaat uit van de volgende assumpties (Stanovich, 1986): 1) 'Word-calling' komt voor als geschreven woorden efficiënt, maar zonder begrip, worden omgezet in spraak, 2) en dat is slecht, want 3) dat betekent dat het kind niet begrijpt dat het doel van lezen begrijpen is; 4) kinderen doen dat 'word-calling' omdat ze de verkeerde leesstrategie hebben aangeleerd; 5) het probleem is dat ze te veel uitgaan van fonologische strategieën. Volgens Stanovich blijkt uit onderzoek dat zelfs bij jonge kinderen decoderen automatisch leidt tot semantische activatie, betekenisverlening, *mits* de betekenis van het woord goed in het geheugen is opgeslagen. Stanovich geeft twee mogelijke verklaringen voor het optreden van het fenomeen 'word-calling' bij een kind. De eerste is dat de betreffende woorden niet tot het vocabulaire van een kind behoren. De andere verklaring is dat 'word-calling' een gevolg is van een zwakke decodeervaardigheid. Er treedt geen begrip op omdat de decodeervaardigheid niet voldoende ontwikkeld is: het decoderen kost zoveel aandacht en is zo traag dat de begripsprocessen verstoord raken. Overigens onderscheidt Goodman wel stadia in de verwerving van de leesvaardigheid (zie paragraaf 1.2.1), maar de uitgangspunten die hij hanteert passen duidelijk bij de 'non-stage'-modellen.

In het *tweede paradigma* gaat men er wel van uit dat er kwalitatieve verschillen zijn in het leesproces van beginnende en ervaren lezers. In de loop van de tijd verwerft een lezer nieuwe en efficiëntere manieren om gedrukt materiaal te identificeren. Verschillen in de wijze van woordidentificatie zijn gerelateerd aan de opbouw van orthografische kennis, en niet zozeer aan het verbeteren van syntactische of semantische kennis. Efficiënter en sneller gebruik van maximale orthografische kennis leidt uiteindelijk tot een beter begrip van de tekst. In de resulterende modellen worden duidelijk verschillende stadia onderscheiden, waarbij elk stadium een additionele en vaak efficiëntere weg bevat om woorden te identificeren. De modellen vallend onder dit paradigma zijn vaker gerelateerd aan meer bottom-up modellen van het leesproces. Alle 'stage'-modellen hechten een grote waarde aan het *leren van de grafeem-foneem-koppeling*.

In dit proefschrift sluiten we ons aan bij het laatste paradigma, mede op grond van de bevindingen in paragraaf 1.2 ten aanzien van het proces van technisch lezen. In tabel 1.4 geven we een overzicht van een aantal belangrijke modellen voor de verwerving van leesvaardigheid, waarbij uitgegaan wordt van kwalitatief verschillende stadia in die ontwikkeling. De indeling in drie niveaus in de tabel is ontleend aan Juel (1991).

Tabel 1.4: Overzicht 'stage'-modellen voor de verwerving van de leesvaardigheid

modellen:	Stadia in verwerving van leesvaardigheid (indeling ontleend aan Juel, 1991)		
	1 selective cue	2 spelling-sound	3 automatic
Gibson (1972, 1976)	1 learning to differentiate graphic symbols	2 learning to decode letters to sounds	3 using progressive higher order units of structure
Chall (1983)	0 pre-reading	1 initial reading /decoding	2 confirmation, fluency, ungluing from print
Gough & Hillinger (1980)	1 paired-associate	2 cryptanalysis/cipher stage	
Mason (1980)	1 context dependency	2 visual recognition	3 letter-sound analysis
Ehri & Wilce (1985, 1987)	1 pre-reader	2 phonetic cue/ novice reader	3 cipher/veteran reader
Ehri, (1991)	1a logographic phase	1b transitional phase	2 alphabetic phase 3 orthographic phase

Grofweg veronderstelt men in de diverse modellen dezelfde stadia in de verwerving van de leesvaardigheid. Het *eerste*, 'selective cue', stadium betreft het identificeren van woorden op grond van enkele willekeurige kenmerken. Het lezen van woorden gebeurt niet op basis van de overeenkomst tussen letters en klanken, maar op basis van een arbitrair kenmerk van een woord, bijvoorbeeld een beginletter. Deze manier van lezen van woorden heeft als probleem dat het steeds moeilijker wordt om de woorden te onthouden, aangezien het steeds moeilijker is om onderscheidende kenmerken te vinden en de associaties tussen kenmerken en woorden arbitrair en onsystematisch zijn (Scott & Ehri, 1990). Mason (1980), Ehri en Wilce (1985, 1987) en Ehri (1991) onderscheiden in hun modellen twee fasen binnen dit stadium, namelijk een fase waarin opvallende kenmerken van een woord worden gebruikt en een overgangsfase waarin de lezer partiële kennis over de letter-klank-koppeling gebruikt.

In *het tweede*, 'spelling-sound', stadium identificeert de lezer de woorden door grafische informatie maximaal te gebruiken, ofwel door expliciet gebruik te maken van de letter-klank-koppeling. Het kind ontdekt dat er een systematische relatie bestaat tussen letters en klanken.

In *het derde*, 'automatic', stadium wordt het decoderen geautomatiseerd. Niet in alle modellen wordt deze fase als kwalitatief verschillend van de voorafgaande fase gezien. We zien hier de discussie terug uit paragraaf 1.2.1 over de perceptuele eenheid van verwerking bij woordherkenning. Sommigen (Gough & Hillinger, 1980; Mason, 1980) zien het automatiseren als het automatisch en snel gebruiken van de kennis over letter-klank-koppelingen. Anderen (Gibson, 1972, 1976; Chall, 1983; Ehri, 1991; Ehri & Wilce, 1985, 1987) geven aan dat uiteindelijk hogere orde structuren (woorden, spellingpatronen, delen van woorden) worden verwerkt gedurende het lezen, zonder dat er een expliciete

fonologische vertaling van het woord plaatsvindt. In deze modellen veronderstelt men dus nog een kwalitatief ander stadium, waarin het kind gebruik maakt van de orthografische kennis, van spellingpatronen bij het lezen van woorden (Ehri, 1991). Ehri geeft echter wel aan dat het fonologisch decoderen aan de basis heeft gestaan van het opslaan van woorden en woorddelen in het geheugen, hetgeen een belangrijk verschil is met de manier waarop 'sight words' in de 'selective cue'-fase worden opgeslagen in het geheugen, namelijk op grond van puur visuele kenmerken (zie Scott & Ehri, 1990). Aansluitend op de bevindingen van Ehri en Wilce (1979, 1983) en de ideeën omtrent het automatiseren van de technische leesvaardigheid (zie paragraaf 1.2.4) willen we binnen dit derde stadium nog een onderscheid maken tussen het 'automatisch' herkennen van woorden en het 'snel' herkennen van woorden.

De fasen, zoals in tabel 1.4 weergegeven, betreffen feitelijk alleen het verwerven van de (nauwkeurigheid en snelheid van de) decodeervaardigheid, het technisch lezen. Chall (1983) gaat verder in haar beschrijving van de leesontwikkeling. De eerste fasen beschouwt ze als 'leren om te lezen', waarbij het vooral gaat om de relatie tussen geschreven en gesproken taal ('to master print'). Na haar tweede fase onderscheidt ze fase 3: 'Reading for learning the new: a first step'. Deze fase, en ook de volgende fasen, betreffen vooral 'lezen om te leren', waarbij de relatie tussen geschreven taal en de betekenis centraal staat ('to master ideas'). We zouden kunnen zeggen dat vanaf fase 3 de overstap wordt gemaakt van technisch lezen naar begrijpend en studerend lezen. De fasen 1 tot en met 3 situeert Chall in het basisonderwijs. De volgende fasen bereikt een leerling pas halverwege het voortgezet onderwijs (fase 4: 'Multiple viewpoints', waarin een leerling leert om te gaan met meerdere standpunten) en daarna (fase 5: 'Construction and reconstruction', waarin de lezer een eigen visie/waarheid ontwikkelt uit de visies/waarheden van anderen). Gezien het onderwerp van dit proefschrift zullen we ons verder beperken tot de stadia betreffende het technisch lezen.

Diverse onderzoeken leveren bewijsmateriaal op dat er inderdaad kwalitatief verschillende stadia bestaan in de verwerving van de leesvaardigheid. Experimenten van Gough (zie Juel, 1991) ondersteunen de aanname dat kinderen starten met het herkennen van woorden op basis van de associatie met één kenmerk dat een woord van andere woorden onderscheidt. Andere onderzoeken (Gough, Juel & Roper-Schneider, 1983; Juel, Griffith & Gough, 1985) laten zien dat de leesfouten van kinderen verschillen afhankelijk van het stadium waarin ze lezen. Leesfouten van leerlingen in de 'selective-cue' fase betreffen vaak woorden die een of meer letters gemeenschappelijk hebben met het eigenlijke woord. Leerlingen die in de 'spelling-sound' fase zitten maken minder lees- en spellingfouten dan leerlingen die nog geen goede kennis van grafeem-foneem-koppelingen hebben. De leesfouten die ze maken betreffen vaak nonsens-woorden als gevolg van een niet-succesvolle poging tot verklanken; de spellingfouten betreffen veelal homofonen (gelijkklinkende woorden) van het eigenlijke woord. Kinderen die leren decoderen passen deze vaardigheid ook toe op woorden die ze nog niet kennen of leren. De decodeerstrategie leidt ertoe dat de uitgesproken woorden vaak meer lijken op de geschreven woorden, ook al zijn het soms non-woorden (Ehri, 1991).

Die kwalitatief verschillende stadia zouden uiteraard een artefact kunnen zijn van het onderwijs, waarin het decoderen via de klankcode benadrukt wordt. Volgens de 'non-stage'-modellen onderbreekt het onderwijs de natuurlijke ontwikkeling van geletterdheid, waardoor er een onnatuurlijk stadium gecreëerd wordt waarin het kind lezen ziet als het uitspreken van woorden (zie het eerder genoemde fenomeen 'word-calling') en niet als het achterhalen van de betekenis. Uit onderzoek naar de verwerving van de leesvaardigheid bij kinderen die geen formele instructie ontvingen, bleek echter dat ook deze kinderen in ieder geval twee stadia doorlopen bij het leren lezen: de 'selective cue' fase en de 'spelling-sound' fase (zie Juel, 1991).

Het leren van de grafeem-foneem-koppeling blijkt zeer belangrijk te zijn voor een goede verwerving van de leesvaardigheid. Het belangrijkste verschil tussen goede en zwakke lezers is gelegen in de snelle vaardigheid van de eersten om kennis over de relatie tussen fonemen en grafemen te gebruiken bij woordherkenning (Assink, Lam & Knuijt, 1995; Juel, 1991; Perfetti, 1984, 1985; Seegers, 1985; Stanovich, 1991). Eerder hebben we al aangegeven dat er weinig bewijs is dat het verschil tussen goede lezers en zwakke lezers gelegen is in het feit dat goede lezers bij woordherkenning beter gebruik maken van de context dan zwakke lezers (zie paragraaf 1.2.3). Het voordeel van een alfabetische taal is de overeenkomst tussen fonemen en grafemen en deze relatie is een veel betrouwbaarder en belangrijker informatiebron voor het herkennen van woorden dan de context (Juel, 1991; Perfetti, 1984, 1985).

Het zo vroeg mogelijk verwerven van de decodeervaardigheid blijkt ook een belangrijke voorspeller te zijn voor de latere vaardigheid in begrijpend lezen (Calfée & Piontkowsky, 1981; Juel, 1991; Perfetti, 1985; Stanovich, 1986). Hierop komen we in paragraaf 1.4 nog terug.

1.3.2 Onderwijs in technisch lezen

In de vorige paragraaf hebben we een drietal fasen onderscheiden in de verwerving van de leesvaardigheid. Uitgaande van deze fasen willen we in deze paragraaf beschrijven hoe in het Nederlandse basisonderwijs daarbij aangesloten wordt. We zullen tevens de (mogelijke) plaats aangeven van de AVI-toetsen en de indeling van leesstof in AVI-niveaus.

In het Nederlandse basisonderwijs onderscheiden we binnen het leesonderwijs drie fasen: het voorbereidend, het aanvankelijk en het voortgezet lezen. Deze indeling is al lange tijd gangbaar in ons onderwijssysteem (Van den Berg & Te Lintelo, 1977b; Caesar, 1982; Malmquist & Brus, 1974). De drie fasen in het leesonderwijs lopen globaal genomen parallel aan de drie stadia van leesverwerving (tabel 1.4).

Het voorbereidend leesonderwijs

De eerste fase, het voorbereidend leesonderwijs, valt globaal genomen samen met het 'selective cue'-stadium en betreft de groepen 1 en 2 van het basisonderwijs. Vele jaren heeft het begrip leesrijpheid ('reading readiness') centraal gestaan in deze eerste fase van het leesonderwijs, waarbij de nadruk lag op het leren van de zogenaamde leesvoorwaarden (Malmquist & Brus, 1974). De gedachte was dat het leesonderwijs alleen effectief kon zijn als de kinderen beschikten over deze leesvoorwaarden. De nadruk lag daarbij

vooral op de formele technische aspecten van het lezen. In deze situatie is de laatste jaren verandering gekomen als gevolg van de opkomst van het begrip 'ontluikende geletterdheid': de geletterdheid van een kind begint al voordat het kan lezen en schrijven ('emergent literacy'; Sulzby & Teale, 1991). Al voordat kinderen naar de basisschool gaan, leren ze thuis, op de peuterspeelzaal of op de crèche veel over (gesproken en geschreven) taal (Bus, 1992). Vaak leren ze spontaan bepaalde vormen van lezen en schrijven. In de nieuwe methoden voor de kleutergroepen staat dit zelfontdekkend lezen centraal. Men wil daarmee bewust aansluiten bij de ervaringen van kinderen met geschreven en gesproken taal (Verhoeven & Mommers, 1989).

Het leren lezen in een alfabetische taal betekent het ontdekken van het alfabetische principe: elementaire eenheden van het schrift komen overeen met eenheden van spraak. Er staan echter enkele obstakels in de weg voor het ontdekken van het alfabetische principe (Perfetti, 1984, 1985). Ten eerste zijn fonemen, met name voor medeklinkers, abstract. Zo is bijvoorbeeld de /t/ in 'tijd' akoestisch niet precies gelijk aan de /t/ in 'kat'. Ten tweede bestaan er geen unieke lettercodes voor de klinkers. Denk bijvoorbeeld aan het woord 'vergeten', waarbij de letter 'e' voor drie verschillende fonemen staat. De waarde van het alfabetische systeem is echter groot. Op basis van de grafeem-foneem-correspondentieregels kan men namelijk woorden lezen en schrijven die men nog niet eerder gezien heeft.

Gough en Hillinger (1980) geven vier condities waaraan voldaan moet zijn, wil een kind in het 'cipher-stadium' ofwel het 'spelling-sound'-stadium komen. Het kind moet:

- 1 voldoende ervaring hebben met paren van gesproken en geschreven woorden;
- 2 begrijpen dat woorden zijn samengesteld uit letters (alfabetisch begrip);
- 3 zich ervan bewust zijn dat woorden zijn samengesteld uit betekenisloze klanken, zogenaamde fonemen (fonologisch bewustzijn);
- 4 begrijpen dat het schrift geëncodeerde spraak is en proberen die code te kraken (cryptanalytic intent).

Het fonologisch bewustzijn is van deze condities het meest onnatuurlijk en het moeilijkst te verwerven. Een bepaalde mate van fonologisch bewustzijn blijkt noodzakelijk (maar niet voldoende) te zijn om succesvol een alfabetische taal te leren (Juel, 1988, 1991; Juel, Griffith & Gough, 1986; Mommers, 1990; Perfetti, 1985; Tunmer & Nesdale, 1985). Veel kinderen zullen het alfabetische principe van onze taal niet spontaan leren doorzien (Perfetti, 1985; Verhoeven & Mommers, 1989). Dit brengt ons in de fase van het aanvankelijk leesonderwijs waarin met name aandacht wordt besteed aan het leren omzetten van geschreven taal in gesproken taal: het leren decoderen (Caesar, 1982; Verhoeven, 1992a).

Aanvankelijk leesonderwijs

Het aanvankelijk leesonderwijs, dat over het algemeen start in groep 3 van het basisonderwijs, verloopt grotendeels parallel aan het 'spelling-sound'-stadium. In deze fase van het onderwijs is er op de eerste plaats aandacht voor het leren van de grafeem-foneemkoppeling, ook wel de elementaire leeshandeling genoemd (Van Dongen, 1984). Op de tweede plaats is er aandacht voor het 'automatiseren' van deze leeshandeling.

De aandacht voor de elementaire leeshandeling zien we duidelijk terug in de methoden voor aanvankelijk lezen, zij het op een verschillende manier. Bijvoorbeeld in 'Veilig leren lezen' (Caesar, 1982; Mommers, 1978) de methode die tot in de jaren tachtig door circa 85% van de scholen werd gebruikt, worden zowel structureerwoorden (globaalwoorden) als de letter-klank-koppeling geleerd, om zo een gevarieerd gebruik van woordherkenningsstrategieën te stimuleren. Men start in de methode met de structureerwoorden. In de methode 'Leeslijn' (De Baar, 1992) daarentegen staat in de eerste fase van het aanvankelijk lezen de indirecte woordherkenning via volledige verklanking centraal, met veel aandacht voor losse letters en lettercombinaties. Pas in een later stadium worden enkele klankzuivere woorden geïntroduceerd, en besteedt men aandacht aan andere woordherkenningsstrategieën (Hol, De Haan & Kok, 1995).

In de diverse methoden zien we tevens een opbouw in de moeilijkheidsgraad van de woorden die achtereenvolgens worden geleerd. De eerste woorden die een kind leert lezen zijn eenvoudige eenlettergrepige woorden van de vorm 'mk', 'mkm' en 'km' (m=medeklinker, k=klinker). Daarna leren ze achtereenvolgens andere eenlettergrepige woorden (bijvoorbeeld mmkm, mkmm, mmkmm, enzovoorts), eenvoudige tweelettergrepige woorden, tweelettergrepige woorden met spellingmoeilijkheden, en tot slot drie- en meerlettergrepige woorden (zie ook Ellerman & Spaaij, 1986).

Wanneer de elementaire leeshandeling wordt beheerst, wordt de aandacht verlegd naar het automatiseren van het technisch lezen: het herkennen van woorden zonder expliciete aandacht voor de letter-klank-koppeling. We zien hier een duidelijke parallel met de bevindingen van Ehri en Wilce (1979, 1983): na de nauwkeurigheidsfase waarin kinderen veel aandacht besteden aan het verwerken van letters en aan de letter-klank-koppeling, gaat de ontwikkeling verder naar de *automatische* fase, waarin ze woorden herkennen zonder aandacht voor de afzonderlijke letters en letter-klank-relaties.

Onderzoek naar de effecten van diverse instructiemethoden op de leesvaardigheid van leerlingen bevestigt de opvatting dat het leren van de letter-klank-koppeling belangrijk is voor het leren lezen. Programma's waarin het leren van deze koppeling centraal staat ('code emphasis') leiden tot betere prestaties van leerlingen op woordherkenning en spelling (Chall, 1967; Johnson & Baumann, 1984; Perfetti, 1985) dan programma's waarin de betekenis centraal staat (zogenaamde 'whole language' of 'language experience methods'). Dit geldt met name voor die programma's waarin deze regels op een deductieve systematische manier worden aangeleerd, werkend van deel-naar-geheel (synthetisch) door het leren samenvoegen van klanken tot woorden, en veel minder voor programma's waarin gewerkt wordt met een inductieve, van-geheel-naar-deel (analytische) strategie (Barr, 1984; Beck, 1981; Johnson & Baumann, 1984).

Vergelijkbare resultaten zijn gevonden in onderzoeken naar Nederlandse methoden voor aanvankelijk lezen (Bus, 1984, 1990; Reitsma, 1990). Daarbij werd tevens geconstateerd dat het belangrijk is om kinderen verschillende woordherkenningsstrategieën te onderwijzen, dus naast de fonologische (indirecte) strategie ook de visueel-orthografische (directe) strategie. Kinderen blijken al vrij snel deze technieken door elkaar te gebruiken bij woordherkenning (Reitsma, 1983).

De resultaten van de verschillende onderzoeken wijzen er niet op dat methoden gericht op het leren van de grafeem-foneem-koppeling leiden tot slechtere resultaten voor het

begrijpen van teksten (Beck, 1981; Johnson & Baumann, 1984; Perfetti, 1985), zoals verondersteld wordt door tegenstanders van deze methoden. Immers vanuit de top-down visie op lezen zou een nadruk op het leren van de grafeem-foneem-koppeling leiden tot verkeerde leesstrategieën en een verkeerde opvatting over lezen bij kinderen (Smith, 1973; zie ook het eerder genoemde fenomeen 'word-calling').

De laatste jaren is er enige verschuiving waarneembaar in de methoden voor aanvankelijk lezen in de richting van zelfontdekkend leren en functioneel leesonderwijs. De overeenkomst tussen de eerder genoemde methoden 'Veilig leren lezen' en 'Leeslijn' is dat vooral uitgegaan wordt van de bottom-up visie van lezen met een nadruk op het leren decoderen. Hier tegenover staat de opvatting dat het leesonderwijs de leerling als uitgangspunt moet nemen. Met name in de traditionele vernieuwingsscholen wordt vanuit dit uitgangspunt gepleit voor het leren lezen volgens de natuurlijke methode, waarin kinderen in een eigen tempo ontdekken hoe het schriftsysteem werkt (Van der Geest & Jorna, 1992; Kurvers & Mooren, 1991). Deze opvatting sluit nauw aan bij de eerder genoemde top-down benadering dat het leren lezen een natuurlijk proces zou zijn, dat parallel verloopt aan de taalontwikkeling (Goodman, 1976a).

Mede onder invloed van de eerder genoemde opvattingen ten aanzien van 'ontluikende geletterdheid' heeft men in de diverse nieuwe methoden (Leeslijn, De Leessleutel, De Leesbus, Balans en Vernieuwd Veilig Leren Lezen) geprobeerd om klassikale kant-en-klare instructie te combineren met het zelfontdekkend lezen (Verhoeven & Mommers, 1989). Van de 'natuurlijke methode' heeft men vooral het principe overgenomen om meer bewust aan te sluiten bij de ervaringen die de kinderen al hebben opgedaan met geschreven en gesproken taal. Men probeert veelal de positieve aspecten van de natuurlijke methoden te combineren met de bewezen voordelen van meer gestructureerde methoden. Het zelfontdekkend lezen staat in alle methoden centraal in het kleuteronderwijs (voorbereidend leesonderwijs), maar in de meeste methoden kiest men daarna (in het aanvankelijke leesonderwijs) voor gerichte instructie. In een aantal methoden kan de leerkracht kiezen uit een gestructureerde weg voor de zwakke lezers en een meer zelfontdekkende weg voor de goede lezers (Arts, Fillipiak & Van Laarhoven, 1992; Bus, Both-de Vries, Buntinx & Hardwarsing, 1994; Filippiak & Van Laarhoven, 1992a, 1992b; Fillipiak, Van Laarhoven & Winnubst, 1992; Harskamp, Deinum & Lawant, 1995; Hol, De Haan & Kok, 1995; Kurvers & Mooren, 1991).

De overgang van het 'spelling-sound' stadium naar een stadium waarin de meeste woorden 'automatisch' herkend worden, zonder duidelijke aandacht voor de letter-klank-koppeling, wordt pas bereikt nadat een lezer veel gelezen heeft. Goede kennis van de grafeem-foneem-koppeling is een voorwaarde voor het leren lezen, maar tegelijkertijd dient men niet te lang stil te staan bij het leren van fonologische regels. Zoveel mogelijk tijd dient besteed te worden aan lezen waarin het begrijpen van een tekst en het plezier beleven aan een verhaal centraal staan. Daardoor blijft de belangstelling van kinderen voor het lezen behouden (Juel, 1991). Dit is namelijk belangrijk omdat kinderen veel oefening nodig hebben voor het verkrijgen van een snelle en efficiënte decodeervaardigheid. Naast 'rote drill and practice' (van buiten leren en oefenen) kan in de instructie in decoderen ook structuur en begrip benadrukt worden, bijvoorbeeld aandacht voor de

opbouw van woorden en zinnen, voor orthografie, enzovoorts (Juel, 1991; Lundberg, 1991).

In dit verband lijken Goodman (1976a) en Smith (1973) wel gelijk te hebben met hun opvatting dat teksten rijk aan vocabulaire en context belangrijk zijn voor het leren lezen. Echter niet omdat contextgebruik de belangrijkste leesstrategie is, maar omdat de context gebruikt kan worden om de letter-klank-koppelingen te leren, om onregelmatig gespelde woorden te leren, enzovoorts (Näslund & Samuels, 1992; Nicholson, 1992; Tunmer & Chapman, 1995). Door teksten rijk aan vocabulaire en context te lezen kunnen kinderen de kennis verwerven die (Perfetti, 1985) onder andere noodzakelijk is voor een efficiënte en snelle woordherkenning, voor het verwerven van woordherkenning als modulair proces: kennis op woord-, letter- en foneemniveau en op orthografisch niveau. In de leesmethoden voor aanvankelijk lezen zal daarom een evenwicht gezocht moeten worden tussen enerzijds teksten met een gecontroleerd vocabulaire die nodig zijn voor het kunnen aanleren van het alfabetische principe, en anderzijds teksten die een rijk vocabulaire bevatten en betekenisvol en begrijpelijk zijn voor kinderen (Chall, 1983). In de nieuwe methoden voor aanvankelijk lezen tracht men dan ook aandacht te besteden aan leesbevordering en leesplezier. De AVI-toets en de bijbehorende indeling van boeken naar AVI-niveau kunnen hierbij eveneens een rol spelen. We zullen daarop terugkomen.

Het onderwijs in technisch lezen gaat nog verder na groep 3. Zoals we net al hebben aangegeven is het aanvankelijk lezen toch vooral gericht op de beheersing van de (elementaire) leeshandeling, waarbij nauwkeurigheid op de eerste plaats komt, gevolgd door het 'automatisch' verlopen van de woordherkenning. In het voortgezet technisch leesonderwijs is het van belang de nadruk te leggen op de snelheid van woordherkenning (Ehri & Wilce, 1979, 1983).

Het voortgezet leesonderwijs

Vanaf groep 4 van het basisonderwijs start het voortgezet lezen, gericht op het leren begrijpen van geschreven materiaal. Binnen het voortgezet lezen wordt een nader onderscheid gemaakt in: voortgezet technisch lezen, begrijpend lezen, studerend lezen, leesbevordering en vrij lezen (Aarnoutse, 1992). Een vergelijkbare indeling zagen we reeds in tabel 1.1, bij het bespreken van AVI. Gezien het feit dat de AVI-toets en de AVI-niveau-indeling bedoeld zijn voor het *technisch* lezen, zullen we ons hiertoe verder beperken. De doelstellingen voor het *voortgezet technisch lezen* betreffen veelal enerzijds het snel en efficiënt kunnen herkennen van woorden en woordgroepen, en anderzijds het correct kunnen verklanken van een tekst ter stimulering van het begrijpend lezen (gebruikmakend van leestekens, met een melodieuze accent, etcetera; zie Aarnoutse, 1992). De fase van het voortgezet technisch lezen valt naar onze mening grotendeels samen met het 'automatic' stadium uit tabel 1.4, waarbij de nadruk ligt op het vergroten van de snelheid van decoderen.

De overgang naar 'automatisch' en vervolgens naar 'snel' decoderen zal echter niet voor alle woorden tegelijkertijd plaatsvinden, maar dit zal geleidelijk gebeuren. De gemakkelijkste woorden en de woorden die het eerst geoefend worden zullen het eerst automatisch en vervolgens snel gedecodeerd worden.

Deze aanname dat het proces van efficiënt en snel decoderen voor verschillende woorden zich in een verschillende tijdperiode ontwikkelt, wordt bevestigd door onderzoek. Verschillende woordkenmerken (lengte, frequentie, lexicale status) beïnvloeden de leessnelheid (Van Bon & Schreuder, 1986; Ehri & Wilce, 1983; Just & Carpenter, 1987; Rayner & Pollatsek, 1989; Reitsma, 1983): hoogfrequente woorden worden sneller gelezen dan laagfrequente, korte woorden sneller dan lange, enzovoorts. Ook worden korte, orthografisch regelmatige woorden sneller geleerd dan lange, orthografisch regelmatige, en worden deze laatste weer sneller geleerd dan orthografisch onregelmatige woorden (Bus, 1990). De leessnelheden verschillen duidelijk tussen goede en zwakke lezers: hoe minder frequent de woorden of hoe langer de woorden, hoe groter de verschillen in leestijden van goede en zwakke lezers (Perfetti, Finger & Hogaboam, 1978; Seegers, 1985). Deze verschillen worden wel kleiner in de loop van de tijd. Dit betekent dat we bij het meten van de technische leesvaardigheid rekening moeten houden met de moeilijkheidsgraad van de woorden en de volgorde waarin leerlingen deze woorden leren in het onderwijs (zie ook bij het aanvankelijk leesonderwijs).

Snelheid is een van de beste indicatoren voor het geautomatiseerd zijn van woordherkenning (Nathan & Stanovich, 1991). De snelheid van woordherkenning wordt verondersteld de toegankelijkheid van de lexicale representaties weer te geven (Perfetti, 1984). De toegankelijkheid van een woord is volgens Perfetti afhankelijk van de kwaliteit van de representatie ervan in het geheugen. Die mentale representaties van woorden omvatten informatie op woord-, letter-, foneem- en orthografisch niveau. De nauwkeurigheid van die representaties, en wellicht hun redundantie en hun onderlinge verbondenheid, maken een snelle woordherkenning mogelijk. In zijn visie dient daarom via oefening de kennis van woorden en van orthografische patronen toe te nemen om de leesefficiëntie te vergroten. Perfetti (1985) geeft aan dat die snelheid niet zozeer bevorderd wordt door alleen te trainen op het snel lezen van woorden, maar door veel oefening, ook of wellicht juist met teksten, waardoor de kennis over de woorden en orthografie vergroot wordt. Deze ideeën stemmen overeen met de in paragraaf 1.2.4 beschreven opvatting dat voor het automatiseren van het proces van woordherkenning de lezer een kennisbestand moet opbouwen door veel te oefenen (Näslund & Samuels, 1992; Stanovich, 1991). Voor leesinstructie betekent deze opvatting dat instructie die gericht is op het bevorderen van woordkennis en orthografische kennis beter is dan instructie die dat niet doet, en dat het trainen op alleen snelheid niet kan compenseren voor het niet-trainen van kennis. De woordherkenningsvaardigheid kan niet alleen verbeterd worden door het oefenen met woorden (Reitsma, 1988), maar bijvoorbeeld ook door het herhaalde malen lezen van dezelfde teksten (Samuels, 1979) en door veel leeservaring.

Het belang van veel leeservaring en oefening voor het leren lezen wordt bevestigd in diverse onderzoeken naar de effecten van 'print exposure' op woordherkenning, leesvaardigheid en spellingvaardigheid (Cipielewski & Stanovich, 1992; Cunningham & Stanovich, 1990, 1991; Stanovich & West, 1989). Zoals we al eerder hebben opgemerkt, is (een bepaalde mate van) fonologisch bewustzijn een noodzakelijke voorwaarde voor de ontwikkeling van een efficiënte woordherkenningsvaardigheid, maar geen voldoende voorwaarde (Juel, Griffith & Gough, 1986; Tunmer & Nesdale, 1985). Met name voor het her-

kennen van onregelmatig gespelde woorden zijn orthografische verwerkingsvaardigheden (de vaardigheid om orthografische representaties te vormen, op te slaan en op te roepen) belangrijk, naast de fonologische codeervaardigheid (Ehri, 1991; Reitsma, 1983). Voor de ontwikkeling van zowel de fonologische codeervaardigheid als de orthografische verwerkingsprocessen (waarbij ook de fonologische codeervaardigheid een rol speelt) blijkt de hoeveelheid leeservaring (print exposure) van belang. Overigens lijkt er tussen de decodeervaardigheid en de leeservaring een wederzijdse beïnvloeding te bestaan: een goede decodeervaardigheid leidt tot plezier in lezen en tot veel lezen, en veel lezen leidt vervolgens weer tot een verbetering van de leesvaardigheid. Daarnaast heeft leeservaring ook een belangrijke invloed op het begrijpend lezen en de woordenschatkennis van leerlingen.

De praktijk van het voortgezet technisch leesonderwijs in Nederland bestaat uit enerzijds instructie en oefening in woordidentificatie en anderzijds uit het veel lezen van eenvoudige teksten of boekjes (Aarnoutse, 1992). Het eerste aspect betreft het bevorderen van het gebruiken van verschillende woordidentificatietechnieken door aandacht voor spellingpatronen, letterclusters en de morfologische structuur van woorden. Het tweede aspect is vooral gericht op het opvoeren van het leestempo via veel leeservaring. Dit gebeurt zowel door stillezen als door hardop lezen. En een vaak gebruikte methode daarbij is het zogenaamde niveaulezen, één van de uitwerkingen van het AVI-innovatieproject.

De rol van AVI in het technisch leesonderwijs

Op het moment dat de leerlingen de elementaire leeshandeling beheersen en een redelijk aantal eenvoudige woorden 'automatisch' kunnen lezen, wordt in veel scholen gestart met het lezen op AVI-niveau. Door leerlingen teksten te laten lezen op een niveau dat ze redelijk beheersen, kunnen ze hun woordherkenningsnelheid verder opvoeren (Beck, 1981; Chall, 1983; Samuels, 1979).

In de oorspronkelijke AVI-publicatie (Van den Berg & Te Lintelo, 1977b) werd het lezen in vorderingengroepen (ook wel niveaulezen genoemd) uitgewerkt als een manier om de technische leesvaardigheid van kinderen te vergroten, rekening houdend met de individuele verschillen tussen kinderen. De bedoeling van het lezen in vorderingengroepen was dat kinderen van eenzelfde AVI-niveau gezamenlijk lezen in groepjes van maximaal 4 kinderen. Om de beurt dienden de kinderen hardop voor te lezen uit een boekje van het betreffende AVI-niveau. Het valt buiten het bestek van dit hoofdstuk om nader in te gaan op het lezen in vorderingengroepen, maar we willen wel wijzen op wat volgens ons voor de AVI-toets van belang is: het aan kinderen kunnen aanbieden van leesstof die afgestemd is op hun technische leesvaardigheid. Belangrijk voor het bevorderen van het technisch lezen is dat, naast een goede instructie, volop geoefend wordt met teksten die niet te moeilijk of te gemakkelijk zijn. De AVI-toets en de bijbehorende niveauindeling van boeken zijn bedoeld als hulpmiddel bij de selectie van dergelijke teksten. Aangenomen wordt dat beheersing van een bepaald AVI-niveau betekent dat een leerling teksten van dat niveau zodanig efficiënt en snel leest dat hij deze zelfstandig en met plezier kan lezen. In hoofdstuk 2, wanneer we de constructie van de AVI-toetskaarten bespreken, gaan we nader in op de opbouw van de AVI-niveaus en de aansluiting van die niveaus op het leesonderwijs.

Wat kunnen we nu leren uit bovenstaande bevindingen als we een toets voor technisch lezen willen maken? De belangrijkste aspecten waarmee we rekening zullen moeten houden zijn ten eerste het bestaan van de verschillende stadia in de leesontwikkeling, ten tweede de snelheid als beste indicator voor de mate waarin de technische leesvaardigheid geautomatiseerd is en ten derde de afstemming op het technisch leesonderwijs, aangezien woorden die het eerst geleerd worden ook het eerst 'automatisch' en snel worden gedecodeerd.

1.4 Technisch lezen, begrijpend lezen en woordenschat

We zijn in paragraaf 1.2 begonnen met de opmerking dat woordherkenning de fundering is van het leesproces, maar dat het uiteindelijke doel van lezen begrijpen is. We willen in deze paragraaf aandacht besteden aan de relatie tussen technisch lezen en begrijpend lezen. Met het oog op het bepalen van de begripsvaliditeit van een toets voor de technische leesvaardigheid is het belangrijk deze relatie wat verder uit te werken. Immers een toets voor technisch lezen moet duidelijk iets anders meten dan een toets voor begrijpend lezen. Maar tegelijkertijd dient de samenhang tussen de prestaties op de beide soorten toetsen overeen te komen met de veronderstelde relatie tussen technisch en begrijpend lezen.

Naast begrijpend lezen willen we tevens aandacht besteden aan woordenschat. Woordherkenning en woordenschat zullen op de een of andere manier aan elkaar gerelateerd zijn, maar zijn zeer verschillend. Zo betreft de woordherkenning met name het technisch kunnen lezen van woorden, terwijl de woordenschat betrekking heeft op het kennen en begrijpen van de betekenis(sen) van woorden.

1.4.1 Technisch lezen en begrijpend lezen

Veelal wordt verondersteld dat goed technisch lezen een voorwaarde is voor het goed kunnen begrijpen van een tekst. Zo gaat Perfetti (1985) er in zijn 'verbal efficiency' model vanuit dat het begrip van een tekst niet maximaal kan zijn als het decoderen niet efficiënt en snel gebeurt. In zijn 'lexical access' hypothese geeft hij aan dat efficiënte lexicale toegang, die weinig capaciteit vereist, het mogelijk maakt dat het werkgeheugen zich bezig houdt met het begrijpen van tekst. Inefficiënte toegang werkt interfererend met de begripsprocessen. Iemand die langzaam decodeert 1) leest niet snel genoeg voor het efficiënt verwerken van zinnen, doordat het werkgeheugen slechts korte tijd informatie kan vasthouden, 2) leest minder woorden van een tekst waardoor de kans om nieuwe woorden in de context te leren kleiner is, en 3) leest vaak ook nog onnauwkeurig, waardoor het begrip van de tekst nog verder verstoord wordt (Nicholson, 1986; Perfetti, 1985).

Ook Stanovich (1986) veronderstelt in zijn 'interactive compensatory model' dat leerlingen met een gebrekkige decodeervaardigheid tevens beperkt zullen worden in het begrijpen van de tekst. Leerlingen kunnen het gebrek aan decodeervaardigheid deels compenseren door gebruik te maken van andere kennisbronnen om woorden te herkennen, maar die kennisbronnen vragen meer aandacht dan een goede decodeervaardigheid waardoor minder aandacht beschikbaar is voor begripsprocessen. De sleutel tot het activeren van

verwerkingscapaciteit voor processen van een hogere orde is de efficiënte decodeervaardigheid.

In zijn beschrijving van het zogenaamde Mattheüs-effect geeft Stanovich (1986) een verklaring voor de invloed van de vaardigheid in decoderen (technisch lezen) op het begrijpend lezen. Kinderen die moeilijkheden ervaren met het leren van de letter-klank-koppeling, zullen al vrij snel minder woorden en teksten lezen dan hun leeftijdsgenoten. Bovendien zullen die teksten voor hen al gauw te moeilijk zijn. Het gevolg hiervan is dat de kinderen in een negatieve spiraal terecht komen. Door de gebrekkige decodeervaardigheid lezen ze minder teksten, en de teksten die ze lezen zijn vaak te moeilijk. Daardoor gaan ze lezen steeds minder leuk vinden, hetgeen resulteert in nog minder lezen. En juist dat gebrek aan oefening bij de zwakkere lezers vertraagt de ontwikkeling van een geautomatiseerde en snelle woordherkenning, hetgeen uiteindelijk ook effecten heeft op de woordenschat en de algemene en syntactische kennis van deze leerlingen (zie ook de onderzoeken naar 'print exposure' in paragraaf 1.3.3).

Diverse onderzoeken bevestigen bovenstaande ideeën dat verschillen in technisch lezen gerelateerd zijn aan verschillen in begrijpend lezen, alhoewel het bewijs veelal geleverd wordt met behulp van correlationeel onderzoek. De betere (begrijpend) lezers lezen woorden nauwkeuriger, maar vooral sneller dan de zwakke begrijpend lezers (Curtis, 1980; Lesgold & Curtis, 1981; Lesgold & Resnick, 1982; Nathan & Stanovich, 1991; Perfetti, 1985; Seegers 1985; Stanovich, 1980; Stanovich, Cunningham & Feeman, 1984). Hun voordeel in nauwkeurigheid lijkt vooral beperkt te zijn tot laagfrequente woorden en pseudowoorden. Hun voordeel in snelheid lijkt altijd te gelden, maar het meest voor onbekende woorden, pseudowoorden en lange woorden.

Zo vindt Seegers (1985) dat goede en zwakke begrijpend lezers vooral verschillen in snelheid van woordherkenning. Deze verschillen zijn het grootst in leerjaar 3 (groep 5) en zijn wat afgenomen in leerjaar 5 (groep 7), althans voor de in zijn onderzoek gebruikte woorden. Dit laatste schrijft Seegers toe aan het feit dat de herkenningssnelheid van de goede lezers op dat moment een asymptotische waarde heeft bereikt, waardoor de zwakkere lezers als het ware de kans krijgen om de goede in te halen. Een vergelijkbaar 'inhaaleffect' vinden Boland en Mommers (1986): in leerjaar 6 (groep 8) hebben de zit-blijvers de voorsprong van de doorstromers in technisch lezen enigszins ingelopen. Onderzoek naar dyslexie geeft eveneens aan dat een langzame en inefficiënte decodeervaardigheid een belangrijke oorzaak is voor leesproblemen (Assink, 1989, 1995; Lovett, 1987; Manis, 1985; Szezulski & Manis, 1987; Yap & Van der Leij, 1995). Het feit dat de verschillen tussen goede en zwakke lezers het grootst zijn bij het lezen van pseudowoorden, geeft aanleiding te vermoeden dat vooral de kennis en het toepassen van de letter-klank-koppeling een probleem opleveren (Assink, Lam & Knuijt, 1995; Seegers, 1985; Yap & Van der Leij, 1995). Samenvattend kunnen we vaststellen dat onderzoek erop wijst dat wat betreft woordherkenning de *snelheid* eerder dan de *nauwkeurigheid* de belangrijkste factor is voor de ontwikkeling van de (begrijpende) leesvaardigheid, hetgeen overeenkomt met de eerder besproken opvattingen over het automatiseren van de woordherkenning.

Het verband tussen technisch en begrijpend lezen blijkt overigens samen te hangen met de leeservaring en de leesvaardigheid: de correlaties zijn het hoogst in de lagere leerjaren. Bij volwassen lezers speelt de snelheid van woordherkenning nauwelijks een rol bij verschillen in leesbegrip (zie Daneman, 1991). Zo rapporteert Boland (1991) correlaties tussen technisch en begrijpend lezen van 0,64 in leerjaar 2 (groep 4), 0,61 in leerjaar 3 (groep 5) en 0,46 in leerjaar 6 (groep 8). Onderzoek van Aarnoutse, Mommers, Smits en Van Leeuwe (1986) laat eveneens deze dalende trend zien met correlaties van 0,51 aan het begin van leerjaar 2 (groep 4) tot 0,41 aan het begin van leerjaar 4 (groep 6). We zien overigens wel grote verschillen tussen de correlaties in beide onderzoeken, bijvoorbeeld 0,64 versus 0,51 voor groep 4. Een mogelijke verklaring hiervoor is dat de correlaties van Boland betrekking hebben op latente variabelen voor technisch en begrijpend lezen (gecorrigeerd voor onbetrouwbaarheid), die gemeten zijn met meerdere toetsen. In het andere onderzoek hebben de correlaties betrekking op telkens één toets voor technisch lezen (de Eén-Minut-Test) en één toets voor begrijpend lezen (die verschilt per tijdstip).

Longitudinaal onderzoek bevestigt het beeld dat technisch lezen voorwaardelijk is voor goed begrijpend lezen (Aarnoutse & Van Leeuwe, 1988; Aarnoutse, Mommers, Smits en Van Leeuwe, 1986; Boland, 1991; Juel, 1988; Juel, Griffith & Gough, 1986; Mommers, 1987; Mommers, Van Leeuwe, Oud & Janssens, 1986). Kinderen die zwak zijn in technisch lezen, zijn dat over het algemeen ook in begrijpend lezen. Leerlingen die aan het begin van het basisonderwijs zwak zijn in begrijpend lezen, zijn dat ook in de latere leerjaren. Prestaties op het gebied van technisch lezen hebben vooral in het begin van het onderwijs een directe invloed op het leesbegrip. De sterkte van deze invloed neemt af naarmate het onderwijs vordert.

Op grond van het bovenstaande veronderstellen we een causale relatie tussen technisch lezen en begrijpend lezen. Hiermee willen we echter niet zeggen dat technisch lezen ook een voldoende voorwaarde is voor begrijpend lezen. De decodeervaardigheid kan gezien worden als een noodzakelijke voorwaarde voor een goede leesvaardigheid, maar niet als een voldoende (Daneman, 1991). Ook andere factoren spelen een rol. In de zogenaamde 'simple view of reading' (Van den Bos, 1989; Gough & Tunmer, 1986; Hoover & Gough, 1990) gaat men er vanuit dat lezen bestaat uit twee componenten: decoderen en taalbegrip, die beide noodzakelijk zijn voor succesvol lezen, maar geen van beide voldoende. Diverse onderzoeken bevestigen het bestaan van deze twee afzonderlijke componenten (Curtis, 1980; Dymock, 1993; Hoover & Gough, 1990; Juel, 1988; Juel, Griffith & Gough, 1986). Zowel decodeervaardigheid (soms gemeten als nauwkeurigheid en soms als snelheid) als luistervaardigheid (als maat voor taalbegrip) is gecorreleerd met leesbegrip, waarbij in de eerste leerjaren decoderen iets sterker dan luisterbegrip gerelateerd is aan leesbegrip, en in de latere leerjaren luisterbegrip sterker gecorreleerd is met leesbegrip (zie ook Daneman, 1991).

Nederlands onderzoek bevestigt dat het verband tussen technisch en begrijpend lezen niet strikt voorwaardelijk is. Een deel van de leerlingen (circa 10%) is zwak in technisch lezen en tegelijkertijd goed in begrijpend lezen (Aarnoutse et al, 1986; Aarnoutse & Van

Leeuwe, 1986). Deze leerlingen blijken over een surplus aan verbale, ruimtelijke en fonologische bekwaamheden te beschikken.

Hoopgevend voor het onderwijs is de bevinding dat leeservaring kan compenseren voor een beperkte decodeervaardigheid. Cunningham en Stanovich (1991) vonden namelijk dat kinderen met een relatief lage decodeervaardigheid, maar met relatief veel leeservaring net zo goed presteerden op begrijpend leestoetsen als degenen die relatief hoog scoren op decodeervaardigheid en relatief maar weinig leeservaring hebben. Door middel van veel leeservaring kunnen leerlingen hun beperkte woordherkenningsvaardigheid blijkbaar zodanig compenseren dat hun leesbegrip op hetzelfde niveau komt als dat van leerlingen met een goede woordherkenningsvaardigheid.

1.4.2 De relatie van technisch en begrijpend lezen met woordenschat

Begrijpend lezen en woordenschat zijn over het algemeen sterk aan elkaar gerelateerd. De gerapporteerde correlaties zijn vrij hoog: 0,73-0,76 in groep 5 en 6 (Stijnen, 1975) 0,66 en 0,63 in groep 7 en 8 (Aarnoutse, 1987), 0,71 voor 10-jarigen, 0,75 voor 14-jarigen en 0,68 voor 18-jarigen (Thorndike, 1973, zie Daneman, 1991). De hoogte van de correlaties lijkt vrij algemeen te gelden voor verschillende taalgebieden (zie Anderson & Freebody, 1985).

Goede begrijpend lezers beschikken over het algemeen over een grotere woordenschat dan zwakkere, en kennen vaak ook meer betekenissen van woorden (Aarnoutse, 1991; De Groot, 1991). Over de causale relatie tussen begrijpend lezen en woordenschat bestaat echter veel discussie: lezen goede begrijpend lezers beter omdat zij meer woorden kennen, of kunnen goede begrijpend lezers beter de betekenis van woorden afleiden waardoor hun woordenschat groter wordt? Voor beide standpunten valt wat te zeggen.

De eerste visie wordt ook wel de *instrumentele* visie genoemd (Anderson & Freebody, 1985), waarbij men er vanuit gaat dat kennis van woorden het begrijpen van teksten vergemakkelijkt: woordenschat heeft een directe invloed op het tekstbegrip. In veel onderzoek wordt woordenschat gezien als een belangrijk middel voor het optimaliseren van het leesbegrip. Expliciete instructie in de betekenis van woorden blijkt echter weinig effect te hebben op het leesbegrip (zie Beck & McKeown, 1991). Instructie die wel effect heeft blijkt gekenmerkt te worden door drie elementen (Beck, Perfetti & McKeown, 1982; McKeown, Beck, Omanson & Pople, 1985; Mezynski, 1983): een grote hoeveelheid oefening (hetzelfde woord meerdere keren aanbieden), veelzijdige informatie over woorden (zowel definities als contextuele informatie) en het stimuleren van actieve of diepe verwerking (zodat netwerkrelaties worden opgebouwd tussen nieuwe en oude informatie). Blijkbaar is kennis van woorden alleen niet voldoende, maar is het belangrijk dat de lezer verbanden legt tussen de verschillende woordbetekenissen. Dit is in overeenstemming met de huidige opvatting dat de woordenschat geen verzameling losse woorden is, maar een hecht netwerk. Woorden worden beter onthouden als ze meer betekenisrelaties hebben met andere woorden, ofwel als ze stevig verankerd zijn in het netwerk (Verhallen, 1991, 1994; Verhallen & Verhallen, 1994).

De tweede visie, de *verbale bekwaamheidsvisie*, betreft het idee dat woordenschat eenvoudigweg taalvaardigheid reflecteert (Anderson & Freebody, 1985). Woordenschat is sterk gerelateerd aan tekstbegrip omdat de taalvaardigheid ook de mate van tekstbegrip beïnvloedt. Een uitwerking hiervan is de 'Learning from the context'-hypothese van

Sternberg en Powell (zie Daneman, 1991). Volgens hen zijn mensen met een grote woordenschat degenen die de vaardigheid bezitten om de betekenis af te leiden uit gesproken en geschreven taal. Zowel woordenschat als tekstbegrip zijn een uiting van het vermogen om nieuwe informatie te leren of af te leiden uit de (geschreven of gesproken) context.

Uit verschillende onderzoeken blijkt inderdaad (Perfetti, 1985; Stanovich, 1986) dat de goede begrijpend lezers niet alleen over een grotere woordenschat beschikken, maar tevens de betekenis van onbekende woorden beter kunnen afleiden en beter de binnen de context passende betekenis van een woord kunnen afleiden. Dezelfde processen die helpen bij het begrijpen van een tekst, helpen ook bij het afleiden van de betekenis van woorden. Zo geredeneerd zijn verschillen in woordenschat eerder het resultaat van verschillen in leesvaardigheid dan de primaire oorzaak. Hierbij willen we wel twee kanttekeningen plaatsen. Ten eerste wordt het grootste deel van de woordenschat verworven via mondeling taalgebruik. Het leren van woorden uit de *gesproken* context is voor de (eerste) taalverwerving een krachtig proces waardoor veel woorden als vanzelf worden verworven (Verhallen & Verhallen, 1994). Ten tweede laat onderzoek naar het leren van woorden uit de *geschreven* context slechts kleine effecten zien (Beck & McKeown, 1991; Nagy, Herman & Anderson, 1985). Pas nadat een lezer een woord vele malen is tegengekomen in teksten is er een effect te vinden op woordenschat. Teksten bevatten vaak ook weinig aanwijzingen om de betekenis van woorden af kunnen te leiden.

Gezien het bovenstaande denken we dat de relatie tussen woordenschat en begrijpend lezen niet een puur instrumentele relatie zal zijn. Naar onze mening blijven er nog twee mogelijkheden over.

De eerste mogelijkheid is wederzijdse beïnvloeding: de ontwikkeling van de woordenschat vergemakkelijkt het begrijpend lezen, en het lezen zelf leidt weer tot een groei van de woordenschat. Deze wederzijdse beïnvloeding leidt tot een cumulatief voordeel voor de goede begrijpend lezers en een cumulatief nadeel voor de zwakkere: kinderen met een onvoldoende woordenschat, die langzaam en met weinig plezier lezen, lezen minder, met als gevolg een langzamere ontwikkeling van de woordenschat, waardoor de groei in begrijpende leesvaardigheid verder belemmerd wordt. De andere mogelijkheid is de eerder besproken verbale bekwaamheidsvisie: woordenschat en begrijpend lezen zijn beide een uiting van de verbale bekwaamheid.

Beide opties lijken veel op elkaar, waardoor er eigenlijk sprake is van een nuance-verschil. In het eerste geval zijn woordenschat en begrijpend lezen aparte grootheden die elkaar beïnvloeden, in het tweede geval zijn beide een meting van hetzelfde onderliggende concept 'verbale bekwaamheid'.

Tot dusver hebben we alleen gesproken over de relatie tussen begrijpend lezen en woordenschat. Maar welke relatie bestaat er tussen het herkennen van woorden (technisch lezen) en het kennen van woorden (woordenschat)?

In twee Nederlandse studies is onderzoek verricht naar de relatie tussen technisch lezen, begrijpend lezen en woordenschat (Aarnoutse & Van Leeuwe, 1988; Verhoeven & Gillijns, 1994). Maar in geen van beide is de relatie tussen technisch lezen en woordenschat expliciet bestudeerd. In de studies veronderstelde men dat zowel technisch lezen als woor-

denschat het begrijpend lezen beïnvloeden. Verhoeven en Gillijns vonden dat in groep 3 en 4 het technisch lezen een iets betere voorspeller van begrijpend lezen was dan woordenschat, althans voor autochtone leerlingen. Voor allochtone leerlingen was de invloed van de woordenschat relatief wat groter. Aarnoutse en Van Leeuwe (1988) vonden voor de groepen 5, 6 en 8 dat woordenschat de belangrijkste voorspeller was van begrijpend lezen, en technisch lezen het minst belangrijk. Ze vonden overigens een vergelijkbare verschuiving als die we eerder hebben besproken ten aanzien van de relatie luisterbegrip, decoderen en begrijpend lezen: in de lagere leerjaren heeft de decodeervaardigheid de grootste invloed op het begrijpend lezen, in de hogere leerjaren (als de decodeervaardigheid sneller verloopt) wordt de invloed van de woordenschat groter.

Beide onderzoeken geven geen aanknopingspunten voor de richting van de relatie tussen technisch lezen en woordenschat. We gaan er in dit proefschrift van uit dat de woordherkenningsvaardigheid invloed heeft op de woordenschatkennis. Gezien de bevindingen dat woordherkenning nauwelijks of niet beïnvloed wordt door hogere-orde processen (waaronder ook de kennis van woorden valt), lijkt een omgekeerde relatie veel minder plausibel. Voor een positieve invloed van woordherkenningsvaardigheid op de woordenschat kunnen we steun vinden bij de ideeën van Stanovich en Perfetti. In hun visie zullen zwakke technische lezers minder lezen, daardoor minder met teksten in aanraking komen, waardoor ze minder kans krijgen om nieuwe woorden en woordbetekenissen te leren via het lezen. Ook het eerder genoemde onderzoek naar "print exposure" geeft hiervoor steun. Uiteraard moeten we ons wel realiseren dat de woordenschat niet alleen via lezen opgebouwd wordt, maar ook (en waarschijnlijk meer) via de gesproken taal.

Samenvattend kunnen we stellen dat we bij het onderzoek naar de validiteit van de AVI-toets als instrument voor technisch lezen, veronderstellen dat technisch lezen een directe invloed heeft op begrijpend lezen en woordenschat. Daarbij gaan we ervan uit dat woordenschat en begrijpend lezen hetzij elkaar wederzijds beïnvloeden, hetzij beide een uiting zijn van verbale bekwaamheid. Daarnaast verwachten we dat de relatie tussen technisch lezen en begrijpend lezen het sterkst is aan het begin van het leesonderwijs en zwakker zal worden naarmate het leesonderwijs vordert.

1.5 Het meten van technisch lezen

Om een goed beeld te kunnen krijgen van de ontwikkeling in technisch lezen bij leerlingen is het van belang periodiek betrouwbare en valide metingen te verrichten. In deze paragraaf bespreken we de manier waarop met de AVI-toets dit technisch lezen gemeten wordt en wat de verschillen zijn met andere toetsen voor technisch lezen. In paragraaf 1.5.1 starten we met een weergave van de elementen waaraan een toets voor technisch lezen zou moeten voldoen, gerelateerd aan de functie van de toets en rekening houdend met de huidige wetenschappelijke kennis over de verwerving van de technische leesvaardigheid. We zullen daarbij aangeven welke keuzes ten aanzien van de AVI-toetskaarten zijn gemaakt. Vervolgens zullen we in paragraaf 1.5.2 de AVI-toetskaarten vergelijken met andere instrumenten voor technisch lezen.

1.5.1 Eisen voor een toets voor technisch lezen

Aan het begin van dit hoofdstuk hebben we aangegeven dat we, om het technisch lezen te kunnen meten, inzicht moeten hebben in de factoren die een rol spelen in het proces van technisch lezen, de wijze waarop kinderen de technische leesvaardigheid verwerven en de relatie van technisch lezen met leesbegrip. We gaan nu na wat de betekenis is van de verkregen inzichten voor het construeren van een toets voor technisch lezen.

De volgende vijf aspecten krijgen daarbij nadere aandacht:

- de definitie van technisch lezen;
- het type testmateriaal waarvan gebruik gemaakt wordt;
- de scoringswijze van de toets;
- het type respons dat van de leerling vereist wordt;
- de wijze van afname van de toets.

De eventuele keuzes die we maken, hangen uiteraard af van het doel waarvoor de toets bedoeld is. Het doel van de AVI-toets is het vaststellen van het niveau van technisch lezen van kinderen om zo leesstof te kunnen kiezen die aansluit bij hun niveau. We willen met de toets nagaan in welke mate de decodeervaardigheid van de leerling is gevorderd in het derde ('automatic') stadium, waarin woorden zowel 'automatisch' als snel gelezen worden. De toets is bedoeld voor leerlingen van halverwege groep 3 tot en met groep 6 van het reguliere basisonderwijs, maar kan ook voor de hogere groepen gebruikt worden.

We kiezen voor de *definitie van technisch lezen* in de brede betekenis van woordherkenning (zie ook paragraaf 1.2). En we beperken het technisch lezen niet tot het beheersen van de grafeem-foneem koppeling, ofwel tot wat Goodman 'recoding' noemt. Daarmee zouden we hooguit kunnen vaststellen in hoeverre een leerling het tweede 'spelling-sound' stadium beheerst. We gaan er namelijk vanuit dat naarmate de decodeervaardigheid verder geautomatiseerd is, de leerling naast de fonologische strategie ook de visueel-orthografische strategie zal gebruiken om woorden te herkennen.

De definitie van technisch lezen bepaalt deels de keuze voor het *type testmateriaal*: letters, pseudowoorden, woorden, zinnen of teksten. Het is duidelijk dat vanuit onze keuze voor de brede definitie van woordherkenning letters afvallen als toetsmateriaal. Ook pseudowoorden, woorden die geen betekenis hebben maar wel goed uitspreekbaar zijn (Van den Bos & Scheepstra, 1993; Van den Bos, Scheepstra & Lutje Spelberg, 1993), komen niet in aanmerking. Deze woorden kunnen alleen gelezen worden door het toepassen van de grafeem-foneem correspondentieregels. Wanneer we onze toets uit pseudowoorden zouden opbouwen, zouden we expliciet de beheersing van de grafeem-foneem koppeling vaststellen waarmee we feitelijk onze toets beperken tot het meten van het beheersen van het tweede 'spelling-sound' stadium.

De mogelijkheden die overblijven zijn woorden, zinnen en teksten. Het idee achter het gebruiken van zinnen en teksten in plaats van woorden is dat de context een rol kan spelen bij het technisch lezen. We hebben gezien dat de context bij de ervaren lezer niet of nauwelijks (via automatische spreidende activatie) een rol speelt bij woordherkenning. Beginnende lezers gebruiken de context wel bij woordherkenning: voor het leren van de letter-klank-koppeling en voor het leren lezen van onregelmatig gespelde woorden. Naar-

mate de decodeervaardigheid meer geautomatiseerd wordt, neemt de invloed van de context af. Dit wijst erop dat we bij het meten van technisch lezen zouden kunnen volstaan met woorden als testmateriaal.

De nieuwe AVI-toets is, net als de oude toets, opgebouwd uit teksten. Een eerste argument om te kiezen voor teksten is de zogenaamde ecologische validiteit: de testsituatie lijkt meer op de werkelijke leessituatie, waarin men eveneens samenhangende teksten leest. Daar komt nog bij dat de AVI-toets in die vorm breed geaccepteerd is in het onderwijs. Een volgende reden is dat de AVI-toets met teksten als testmateriaal een goede aanvulling is op de bestaande woorden-tests voor technisch lezen. Een laatste reden is dat kinderen, in tegenstelling tot ervaren lezers, wél gebruik maken van de context bij het technisch lezen. Het gebruik van een samenhangende tekst geeft degenen met een minder goed ontwikkelde decodeervaardigheid de mogelijkheid om te compenseren via het gebruik van de context. Naar ons idee zal dit niet snel leiden tot een verkeerde inschatting van die decodeervaardigheid, gezien de vertragende werking van het gebruik van hogere-orde processen bij woordherkenning. Hoe meer een kind gebruik maakt van de context, des te langzamer zal zijn leessnelheid zijn. De invloed van de context zal bij de betere lezers zeer beperkt zijn, alhoewel er mogelijke effecten zijn van de zogenaamde spreidende activatie (zie paragraaf 1.2.3).

De inzichten ten aanzien van de verwerving van de technische leesvaardigheid stellen nog een andere eis aan het testmateriaal, namelijk ten aanzien van de moeilijkheidsgraad. Gebleken is namelijk dat maximale snelheid van decoderen niet voor alle woorden tegelijk wordt bereikt: korte, regelmatig gespelde woorden (vaak ook de woorden die het eerste worden geleerd) worden eerder 'automatisch' en snel gelezen dan lange, onregelmatig gespelde woorden. Dit betekent dat een toets voor technisch lezen woorden van uiteenlopende moeilijkheidsgraad moet bevatten, van eenvoudige eenlettergrepige woorden tot drie- en meerlettergrepige woorden. Vanuit testtheoretisch oogpunt zou men kunnen kiezen voor het gebruik van woorden in plaats van zinnen of teksten, omdat het gemakkelijker is om een reeks woorden te laten oplopen in moeilijkheidsgraad dan een reeks zinnen of teksten. Bij het maken van een zin of tekst moeten moeilijke en makkelijke woorden elkaar enigszins afwisselen om een lopend geheel te krijgen, waardoor het lastiger is om de moeilijkheidsgraad onder controle te krijgen. De keuze voor teksten betekent dat bij de toetsconstructie veel aandacht besteed moet worden aan de moeilijkheidsgraad van de teksten. In hoofdstuk 2 zullen we daarop nader ingaan.

Een volgende keuze bij de toetsconstructie heeft te maken met de *scoringswijze* van de toets. De toets zal zowel de nauwkeurigheid (aantal fouten) als de snelheid (benodigde leestijd) van woordherkenning moeten vaststellen. Voor het bepalen van de technische leesvaardigheid dient de nadruk te liggen op de snelheid, aangezien de nauwkeurigheid waarmee woorden gelezen worden al snel groot is en niet meer verschilt tussen leerlingen. Bevestiging voor deze keuze vinden we in de theorie over technisch lezen als geautomatiseerd proces, waarin duidelijk naar voren komt dat een geautomatiseerde decodeervaardigheid betekent dat een woord nauwkeurig, 'automatisch' en snel herkend wordt. Onderzoek naar de verwerving van het technisch lezen maakt duidelijk dat de ontwikkeling ook in die volgorde verloopt: eerst nauwkeurig, dan automatisch en tot slot

steeds sneller. Snelheid is de belangrijkste indicator voor de mate waarin de technische leesvaardigheid is geautomatiseerd.

Uitgaande van een toets waarmee we willen vaststellen in welke mate de decodeervaardigheid van de leerling geautomatiseerd is, hebben we kort aangegeven uit welk materiaal zo'n toets zou moeten bestaan (minimaal woorden, van uiteenlopende moeilijkheidsgraad) en op basis waarvan vastgesteld wordt in hoeverre het decoderen geautomatiseerd is (nauwkeurigheid en snelheid). We komen nu bij de vraag: wat laten we een kind precies doen om de technische leesvaardigheid vast te stellen? Dit is de vraag naar het *type respons* dat van een kind verwacht wordt.

In de meeste toetsen dient een kind hardop woorden of teksten voor te lezen, waarbij de leestijd en het aantal fouten worden vastgesteld. Een voordeel van het hardop lezen is dat nagegaan kan worden welke fouten een leerling precies maakt. Het is echter de vraag of het hardop voorlezen wel een valide manier is om de technische leesvaardigheid vast te stellen. Immers, het hardop lezen van woorden vraagt meer dan alleen het herkennen van het woord: de leerling moet het gelezen woord vervolgens ook nog omzetten in spraakklanken. Toevallige versprekingen en mogelijke spraakproblemen zullen de resultaten op de toets (aantal fouten en benodigde leestijd) beïnvloeden. Een andere mogelijkheid om de decodeervaardigheid vast te stellen is via lexicale decisie: de leerling moet bij elk woord in een lijst van woorden en pseudowoorden, aangeven of het woord bestaat of niet. Uit onderzoek van Van Bon, Tooren en Van Eekelen (1995) blijkt dat de correlatie tussen het hardop lezen en de lexicale decisie vrij hoog is, hetgeen zou wijzen op een grote overeenkomst in taakuitvoering. Wel vonden ze dat een klein deel van de leerlingen bij herhaling slechter is in hardop lezen dan in lexicale decisie. Een uitspraakprobleem zou daaraan ten grondslag kunnen liggen.

Voor de nieuwe AVI-toets hebben we, analoog aan de oude AVI-toets, gekozen voor hardop lezen, ondanks de beperking die daaraan kleefte. Dit betekent wel dat de AVI-toets niet geschikt is voor leerlingen met een structureel spraakprobleem. Voor leerlingen zonder spraakproblemen verwachten we dat het hardop lezen een goede indicatie geeft van de technische leesvaardigheid. De extra informatie (voor eventuele diagnostische doeleinden) die beschikbaar komt over het type fouten dat een leerling maakt, weegt ons inziens voor deze leerlingen op tegen het nadeel van de invloed van spraak.

Het zal duidelijk zijn dat de keuze van het type respons gevolgen heeft voor de *wijze waarop de toets wordt afgenomen*. De keuze voor hardop lezen betekent dat de afname individueel dient te geschieden.

1.5.2 De AVI-toets vergeleken met andere toetsen voor technisch lezen

In de vorige paragraaf hebben we vijf aspecten besproken die van belang zijn voor het construeren van een toets voor technisch lezen. In deze paragraaf geven we een overzicht van de verschillende toetsen voor technisch lezen die momenteel in het Nederlandse basisonderwijs gebruikt worden, waarbij we aangeven welke keuzes men bij de verschillende toetsen heeft gemaakt ten aanzien van die vijf punten. Daarnaast willen we de toetsen ook vergelijken op aspecten die betrekking hebben op toetsen in het algemeen, zoals het gebruik (norm- of criteriumgericht) en het doel van de toets (algemene ontwikkeling of diagnostiek).

Wanneer we spreken over het volgen van de ontwikkeling in leesvaardigheid van leerlingen in het onderwijs, is het eerste onderscheid dat we kunnen maken dat tussen *informele en formele toetsen* (Verloop & Van der Schoot, 1995). De leerkrachten in het basis-onderwijs gebruiken veelvuldig *informele evaluatieprocedures*, via de dagelijkse interactie met leerlingen. Daarbij kunnen we denken aan observaties van leerlingen tijdens de les, aan het gebruik van opdrachten, controletaken, mondelinge of schriftelijke overhoringen, enzovoorts. Op deze wijze verzamelt de leerkracht informatie over ecologisch valide gedrag, namelijk het gedrag van leerlingen in een normale leessituatie. Daarbij is het belangrijk dat veel verschillende instrumenten en procedures worden gebruikt om zo tot een valide oordeel te komen. Informele toetsen zijn vaak nauw verbonden met de gevolgde leesmethode (Calfee & Hiebert, 1991). Het doel is vaak het opsporen van sterke en zwakke punten bij de leerlingen en het verkrijgen van informatie om het onderwijs te plannen en bij te stellen. Informele toetsen zijn dan ook vaak methode-gebonden. In de Verenigde Staten vormen de zogenaamde '*Informal Reading Inventories*' (IRI's), toetsen gemaakt door de leerkrachten zelf, een veel gebruikte manier van informeel toetsen (Alexander & Heathington, 1988; Farr & Carey, 1986). Dergelijke IRI's bestaan veelal uit een lijst globaalwoorden en stukken tekst van oplopende moeilijkheidsgraad, parallel aan de niveaus in de gebruikte leesmethode. De teksten worden zowel hardop als stil gelezen, en na afloop worden begripsvragen gesteld. Bij het hardop lezen worden de gemaakte fouten en eventueel de benodigde leestijd bijgehouden. Men onderscheidt vier prestatieniveaus: zelfstandig leesniveau (niveau waarop de leerling zelfstandig kan lezen), instructieniveau (niveau waarop leerling kan functioneren met hulp van de leerkracht), frustratieniveau (niveau waarop de leerling te veel fouten maakt en te weinig begrijpt, zelfs met hulp van de leerkracht) en het niveau van luisterbegrip (niveau waarop de leerling voorgelezen materiaal goed kan begrijpen).

Tot de potentieel sterke punten van de IRI's behoren de ecologische validiteit, de 'face' validiteit, de hoge inhoudsvaliditeit, de geringe kosten en het gemak waarmee ze gebruikt en geïnterpreteerd kunnen worden. De kwaliteit van de IRI's is echter sterk afhankelijk van de expertise van de leerkracht in het maken van toetsen. Zwakke punten van IRI's zijn dan ook bijvoorbeeld de onevenwichtigheid in samenstelling, de onbetrouwbaarheid ervan en de moeilijkheid om een adequaat criteriumniveau te bepalen (Alexander & Heathington, 1988).

We willen wijzen op de *overeenkomst in werkwijze* van de AVI-toetskaarten met de eerder genoemde Informal Reading Inventories (IRI's). In beide gevallen worden teksten van oplopende moeilijkheidsgraad voorgelegd aan kinderen, die deze tekst hardop voorlezen; de leestijd en de fouten worden gescoord. Een andere overeenkomst is dat de leessnelheid en het aantal fouten samen gecombineerd bepalen of een leerling een bepaald niveau beheerst. De criteria die bepalen wat een leesfout is en wat niet, komen eveneens overeen (zie paragraaf 1.1.2 en hoofdstuk 2). De IRI's gaan echter verder dan de AVI-toetskaarten: de kinderen lezen de teksten ook stil voor zichzelf en beantwoorden daarna een aantal begripsvragen.

We zullen ons in deze paragraaf verder beperken tot de *formele toetsen*, ook wel gestandaardiseerde toetsen genoemd. Binnen deze categorie kunnen we onderscheid maken tussen *normgerichte toetsen en criteriumtoetsen* (Alexander & Heathington, 1988; Calfee

& Hiebert, 1991; De Corte, Geerligs, Lagerweij, Peters & Vandenbergh, 1981; Farr & Carey, 1986; Johnston, 1984). Met behulp van *normgerichte* toetsen wordt de score van een leerling vergeleken met de scores van een representatieve groep leerlingen. De prestaties van deze representatieve groep vormen het referentiepunt voor de evaluatie van de toetsscore van een leerling. De inhoud van dergelijke toetsen is vaak methode-onafhankelijk en bedoeld voor een brede range van leesvaardigheidsniveaus. Voordelen van normgerichte toetsen zijn: de zorgvuldige ontwikkeling door experts, vastgestelde betrouwbaarheid en validiteit, het kunnen vergelijken van algemene prestatieniveaus, en de bruikbaarheid voor de signalering van zwakke leerlingen. Nadelen zijn onder andere dat het leesgedrag vaak niet overeenkomt met de werkelijkheid, dat het een momentopname betreft onder speciale condities, dat de diagnostische waarde vaak beperkt is en dat zelden suggesties gegeven worden voor eventueel aanvullende instructie.

Criteriumtoetsen worden gebruikt om te bepalen hoe goed een leerling bepaalde doelen of taken beheerst. In veel leesmethoden zijn criteriumtoetsen als voortgangstoets opgenomen om na te gaan in hoeverre leerlingen bepaalde (vaak zeer specifieke) doelen bereikt hebben. Een groot voordeel van criteriumtoetsen betreft de sterke relatie met de instructiedoelen. De toetsen kunnen goede diagnostische informatie geven ten aanzien van deze doelen.

Het referentiepunt van een criteriumtoets wordt gevormd door een absoluut prestatieniveau op een taak. Dit referentiepunt is echter tevens het grootste probleem bij dit type toetsen: wanneer kunnen we spreken van beheersing? Het gekozen criterium is vaak arbitrair (men kiest vaak voor 80% goed) en subjectief, waardoor in de praktijk het onderscheid tussen normgerichte toetsen en criteriumtoetsen vervaagt (Farr & Carey, 1986). Een tweede probleem met criteriumtoetsen is de vraag of de onderliggende veronderstelling dat de leesvaardigheid op te delen is in opeenvolgende doelen en vaardigheden, wel juist is. Toetsen kunnen overigens zowel normgericht als criteriumgericht zijn.

Een volgend onderscheid dat we kunnen maken is dat tussen *algemene toetsen* voor leesvaardigheid en *diagnostische toetsen* (Alexander & Heathington, 1988; Farr & Carey, 1986). Toetsen van de eerste categorie geven algemene informatie over de leesontwikkeling, en zijn vaak normgericht: hoe presteert een leerling vergeleken met andere leerlingen. Dergelijke toetsen kunnen wel een signalerende functie hebben: welke leerlingen raken achterop in hun leesontwikkeling. Diagnostische toetsen geven uitgebreidere informatie over specifieke onderdelen van de leesvaardigheid. Ze geven informatie over de aard van de problemen. Daarbij kan nog onderscheid worden gemaakt tussen diagnostiek op descriptief niveau, toetsen die aangeven welke leerstof problemen oplevert, en op etiologisch niveau, toetsen die aangeven welke onderliggende vaardigheden in onvoldoende mate beheerst worden (De Corte et al, 1981; Verhoeven, 1983; Verloop & Van der Schoot, 1995).

In tabel 1.5 hebben we een aantal toetsen voor het technisch lezen op een rijtje gezet die in het Nederlandse onderwijs gebruikt worden. Naast de indelingen in normgericht/criteriumgericht en algemeen/diagnostisch komen ook de eerder genoemde vijf aspecten aan de orde komen, alsmede de doelgroep in het basisonderwijs waarvoor de toetsen bedoeld zijn. We hebben in het overzicht de niet-methodegebonden formele toetsen opge-

nomen die expliciet bedoeld zijn voor het technisch lezen en momenteel nog verkrijgbaar zijn. Toetsen die leesvoorwaarden meten, zoals tests voor het bepalen van het fonologisch bewustzijn, auditieve en visuele discriminatie, enzovoorts (zie bijvoorbeeld Van Bon, 1984), vallen buiten dit kader. Voor dit overzicht hebben we gebruik gemaakt van overzichtspublikaties over toetsen en tests (Clijsen, Van Bruxvoort, Glaudé, Kooistra & Van der Leij, 1988; Koornneef, Van der Horst & Aukes, 1994; NICL, 1989, 1993). In totaal hebben we 13 toetsen opgenomen in het overzicht, waarbij we de twee Tempo Tests Lezen (woorden en zinnen) respectievelijk de drie toetsen Technisch lezen (1, 2 en 3) in één kolom hebben opgenomen.

Op welke punten verschillen de toetsen opgenomen in tabel 1.5 van elkaar of komen ze juist met elkaar overeen? Het eerste kenmerk waarop de toetsen kunnen verschillen is het gebruik: zijn het *normgerichte toetsen of criteriumtoetsen*? De meeste toetsen zijn normgerichte toetsen: de scores van de leerlingen worden vergeleken met die van een referentiegroep. Deze toetsen vervullen wel een signaleringsfunctie: de score van de leerlingen geeft aan welke leerlingen een achterstand hebben opgelopen ten opzichte van hun leeftijdsgenoten. Drie toetsen zijn duidelijk criteriumtoetsen, namelijk Technisch Lezen 1, 2 en 3 (Verhoeven, 1980a, 1980b; Krom, 1988). Elk van deze toetsen heeft betrekking op een doelstelling die een bepaalde fase van het leesproces afsluit, bijvoorbeeld voor Technisch Lezen 1: 'het kunnen lezen van alle klankzuivere woorden van het type medeklinker-klinker-medeklinker' (Verhoeven, 1983). Voor de drie toetsen geldt dat als criterium voor beheersing 85% goede antwoorden wordt gehanteerd.

Twee toetsen zijn als normgerichte toets én als criteriumtoets gekarakteriseerd: de AVI-toetskaarten (Van den Berg & Te Lintelo, 1977b; Visser, Van Laarhoven & Ter Beek, 1994) en de Diagnostische Toetsbatterij Technische Leesvaardigheid (DTTL; Wassenaar, 1989). Voor de oude en de nieuwe AVI-toetskaarten gelden voor iedere toetskaart criteria voor de leestijd en het aantal fouten die bepalen of een leerling het betreffende AVI-niveau beheerst. Daarnaast biedt de *nieuwe* AVI-toets ook de mogelijkheid om de AVI-niveaus van de leerlingen te vergelijken met die van een referentiegroep (zie hoofdstuk 3). De DTTL bestaat uit twee delen: een niveaubepalend deel met drie normgerichte toetsen, en een diagnostisch deel met vier toetsen gericht op vier specifieke deelvaardigheden van het technisch lezen. Deze laatste kunnen als criteriumtoetsen worden beschouwd, waarbij een beheersingspercentage geldt van minimaal 80%.

De meeste toetsen voor technisch lezen zijn toetsen waarmee we de *algemene ontwikkeling* van de technische leesvaardigheid kunnen vaststellen. Technisch Lezen 1, 2 en 3 zijn daarentegen *diagnostische* toetsen. Twee toetsen zijn bedoeld zowel voor het vaststellen van de algemene ontwikkeling als voor diagnostiek: de Differentiële Zinnenleestest (DZT; Dommerholt, 1982) en de DTTL. Tot slot willen we nog de Drie-Minuten-Toets (DMT; Verhoeven, 1992b) vermelden. Deze toets is zelf geen diagnostische toets, maar bij de toets zijn wel een tweetal aanvullende diagnostische toetsen opgenomen.

Tabel 1.5: Overzicht toetsen voor het technisch lezen

	AVI-toets (KPC; Van den Berg & Te Lintelo 1977b, Visser et al, 1994)	Bodegrav. leestoets (SBD Midden Holland&Rijn- streek, 1992)	Different. zinnen- leestest (Dommer- holt, 1982)	Diagn. Techn. Leesv (Wassenaar, 1989)	Dne- Minuten- Toets (Cito; Verhoeven, 1992b)	Een- Minuut-Test (Brus & Voeten 1973)	Klepel (Van den Bos et al, 1994)	Leesvhtest Leesvht mechan. leesvht (Wiegersma, 1986)	Tempo Test Le- zen zinnen /woorden (De Vos 1992 1980ab; Krom 1988)
I GEBUIK TOETS									
a normgericht	*	*	*	*	*	*	*	*	*
b criteriumtoets	*			*					*
II DOEL TOETS									
a algemene lees- ontwikkeling	*	*	*	*	*	*	*	*	*
b diagnostiek			*	*					*
III DEFINITIE									
a grafeem-foneem koppeling				*			*		
b woordherkenning	*	*	*	*	*	*	*	*	*
IV MATERIAAL									
a pseudowoorden		*		*	*	*	*	*	*
b woorden									*
c zinnen			*						*
d teksten	*								
V SCORINGSWIJZE									
a nauwkeurigheid	*	*	*	*	*	*	*	*	*
b snelheid	*	*	*	*	*	*	*	*	*
VI RESPONSTYPE									
a hardop lezen	*		*	*	*	*	*	*	*
b meerkeuze items		*							*
VII WIJZE VAN AFNAME									
a individueel	*		*	*	*	*	*	*	*
b groepsgewijs		*							*
c tijdslimiet		*	*	deels	*	*	*	*	*
VIII DOELGROEP	3-8	3-8	3-4	3-6	3-8	3-8	3-8	5-8	3-8 3-4

Het derde kenmerk in tabel 1.5 betreft de *definitie van technisch lezen*. Uit het schema kan afgelezen worden dat in de meeste toetsen technisch lezen geoperationaliseerd is vanuit de brede definitie van technisch lezen. In één toets is expliciet uitgegaan van de beperkte definitie van 'recoding', namelijk in de Klepel (Van den Bos, Lutje Spelberg, Scheepstra & De Vries, 1994). Hierbij moeten we wel opmerken dat de Klepel bedoeld is om in samenhang met de Eén-Minuut-Test (EMT; Brus & Voeten, 1973) gebruikt te worden, die wel uitgaat van de brede definitie. De DTTL bevat subtoetsen die de beperkte definitie hanteren en subtoetsen die uitgaan van de brede definitie.

Een volgend onderscheid is dat naar het *type testmateriaal*: zijn de toetsen opgebouwd uit pseudowoorden, woorden, zinnen of teksten? In de toetsen waarmee men expliciet het toepassen van de grafeem-foneem koppeling wil vaststellen is gekozen voor het aanbieden van pseudowoorden. In de meeste andere toetsen worden woorden gebruikt. In de Differentiële Zinnenleestest en de Tempo Test Lezen Zinnen (De Vos, 1992) vormen daarentegen zinnen het basismateriaal. De AVI-toets bevat gehele teksten als testmateriaal. Het idee achter het gebruik van zinnen en teksten in plaats van woorden is dat de context een rol kan spelen bij het technisch lezen. Eerder hebben we gezien dat dit effect beperkt zal zijn. Het lezen van zinnen en teksten komt wel meer overeen met de normale leessituatie.

De meeste toetsen bevatten materiaal van uiteenlopende moeilijkheidsgraad. De toetsen die uit (pseudo)woorden zijn samengesteld zijn veelal zo geconstrueerd dat een kind begint met de meest eenvoudige (pseudo)woorden (eenlettergrepige van het type 'klinker-medeklinker', 'medeklinker-klinker' of 'medeklinker-klinker-medeklinker') en vervolgens steeds moeilijker woorden leest (tot drie- of meerlettergrepige woorden). Hetzelfde geldt voor toetsen die bestaan uit zinnen of teksten: deze zijn eveneens geordend van makkelijk naar moeilijk. Voor de DMT en de DTTL geldt dat deze wel woorden van verschillende moeilijkheidsgraad bevatten, maar dat deze geordend zijn per kaart voor de DMT, respectievelijk per subtest voor de DTTL. Zo bevat bijvoorbeeld bij de DMT kaart 1 alleen woorden van de vormen 'km', 'mk', en 'mkm', kaart 2 de overige eenlettergrepige woorden en kaart 3 de twee-, drie- en vierlettergrepige woorden.

Voor enkele toetsen geldt dat deze slechts materiaal van een beperkte moeilijkheidsgraad bezitten. Technisch Lezen 1 bevat alleen mkm-woorden, Technisch Lezen 2 de overige eenlettergrepige woorden, en Technisch Lezen 3 twee- en drielettergrepige woorden. Ook de moeilijkheidsgraad van het materiaal in de DZT is beperkt.

Het vijfde kenmerk waarop de toetsen van elkaar verschillen is de *scoringswijze*: wordt met de toets *nauwkeurigheid en/of snelheid* van technisch lezen gemeten? In het overzicht hebben we aangegeven dat in de meeste toetsen beide aspecten meegenomen worden bij het bepalen van de toetsscore. Bij de toetsen met een tijdslimiet gebeurt dit doordat de toetsscore bepaald wordt door het aantal goed gelezen woorden in de opgegeven tijdslimiet. Bij de AVI-toetskaarten spelen criteria voor de leestijd en het aantal fouten een rol bij het bepalen van de prestatie van een leerling (zie paragraaf 1.1.2). De Leesvaardigheidstest voor het onderzoek van de mechanische leesvaardigheid (Wiegersma, 1986) en Technisch Lezen 1, 2 en 3 meten echter alleen de nauwkeurigheid van het

technisch lezen. Het niet meenemen van de snelheid van woordherkenning doet naar onze mening niet voldoende recht aan het meten van de technische leesvaardigheid.

De meeste toetsen vereisen als *type respons* het hardop lezen. De Bodegravense leestoets (SBD Midden Holland & Rijnstreek, 1992) en Technisch lezen 1, 2 en 3 bestaan uit meerkeuze items, waarbij een leerling het goede antwoord moet aankruisen.

De *wijze van afname* hangt duidelijk samen met het *type respons*. Toetsen waarbij een leerling hardop moet lezen, dienen individueel te worden afgenomen, de toetsen met meerkeuze items zijn geschikt voor groepsgewijze afname. Daarnaast blijkt dat veel toetsen met een *tijdslimiet* moeten worden afgenomen (veelal tussen 1 en 5 minuten); het zijn zogenaamde speedtests.

Tot slot hebben we in het overzicht aangegeven wat de *doelgroep* is van de toetsen. De meeste toetsen zijn bestemd voor afname in de gehele basisschoolperiode, veelal vanaf het einde van groep 3. De DZT en Technisch lezen 1, 2 en 3 zijn alleen geschikt voor de groepen 3 en eventueel 4 in het basisonderwijs. Dit is verklaarbaar vanuit het eerder besproken gegeven dat het testmateriaal van deze toetsen van een beperkte moeilijkheidsgraad is, waarbij Technisch lezen 1, 2 en 3 ook nog beperkt zijn tot het meten van de nauwkeurigheid.

Wanneer we nogmaals teruggaan naar de keuzes die we in de vorige paragraaf hebben geformuleerd, dan zijn drie daarvan cruciaal gezien het doel van de toets:

- het materiaal moet bestaan uit woorden, zinnen of teksten;
- deze woorden, zinnen of teksten moeten van uiteenlopende moeilijkheidsgraad zijn;
- de toets moet zowel de nauwkeurigheid als de snelheid van woordherkenning meten.

Zes van de beschreven toetsen voldoen aan deze eisen: de AVI, de DTTL, de DMT, de EMT en de twee Tempo-Tests-Lezen. Deze toetsen verschillen wel op andere punten van elkaar. Zo bevatten alle toetsen, behalve AVI en de Tempo-Test-Lezen-Zinnen, woorden als toetsmateriaal. De AVI-toets en één van de subtests van de DTTL zijn de enige toetsen zonder tijdslimiet. De DTTL bevat naast de niveaubepalende subtests ook enige diagnostische toetsen, en bij de DMT horen eveneens enkele diagnostische tests.

We kunnen op basis van de beschreven informatie niet aangeven welke toetsen beter of slechter zijn. We hebben met deze vergelijking alleen willen aangeven welke positie de AVI-toets in het scala aan toetsen inneemt. Alles overziend, kunnen we de AVI-toets karakteriseren als een *criteriumtoets*, die ook *normgericht* gebruikt kan worden, waarmee de *algemene ontwikkeling* van de technische leesvaardigheid in kaart gebracht kan worden. Technisch lezen wordt daarbij opgevat in de *brede betekenis van woordherkenning*. De leerlingen krijgen *individueel teksten van oplopende moeilijkheidsgraad* voorgelegd die ze *hardop* moeten voorlezen. De *snelheid en nauwkeurigheid* waarmee ze dat doen bepalen hun toetsscore. Ondanks het feit dat de inzichten over het technisch lezen wel enigszins zijn veranderd sinds de constructie van de eerste AVI-toets, kunnen we vaststellen dat de opzet ervan voldoet aan de belangrijkste eisen voor een toets ter bepaling van de mate waarin de technische leesvaardigheid geautomatiseerd is. In het volgende hoofdstuk gaan we nader in op de constructie van de nieuwe AVI-toets.

2 DE ONTWIKKELING VAN DE NIEUWE AVI-TOETSKAARTEN

In dit hoofdstuk staat de ontwikkeling van de nieuwe toetskaarten centraal. We beginnen in paragraaf 2.1 met de bespreking van de criteria op basis waarvan de negen AVI-niveaus zijn onderscheiden. In paragraaf 2.2 bespreken we de constructie van de nieuwe toetskaarten. Tot slot worden in paragraaf 2.3 de onderzoeksvragen van het proefschrift nader geconcretiseerd.

2.1 De criteria voor de negen AVI-niveaus

In paragraaf 1.5.1 hebben we reeds aangegeven dat het belangrijk is om de toetskaarten te laten oplopen in moeilijkheidsgraad. Dit omdat de maximale snelheid van decoderen niet voor alle woorden tegelijkertijd bereikt wordt. Korte, regelmatig gespelde woorden en frequent voorkomende woorden worden veelal eerder 'automatisch' en snel herkend dan lange, onregelmatig gespelde woorden en minder frequent voorkomende woorden (Van Bon & Schreuder, 1986; Ehri & Wilce, 1983; Reitsma, 1983; zie ook paragraaf 1.3.2). De indeling in moeilijkheidsgraad van de negen AVI-niveaus is voor een groot deel gebaseerd op de Leesindex A van Brouwer (1963). We zullen daarom eerst nader ingaan op leesbaarheidsformules in het algemeen en de Leesindex A van Brouwer in het bijzonder.

2.1.1 De leesbaarheid van teksten: leesbaarheidsformules

De leesbaarheid van teksten, of beter gezegd het meten van die leesbaarheid, heeft onderzoekers al lange tijd beziggehouden (Klare, 1984). Vanaf de jaren twintig ontwikkelde men de zogenaamde leesbaarheidsformules. Het doel daarvan was om instrumenten te verkrijgen waarmee verschillen in leesbaarheid van teksten voorspeld konden worden. De formule van Vogel en Washburne uit 1928 vormde het prototype voor de meeste formules die daarna verschenen (Klare, 1984). De algemeen gevolgde werkwijze bij het ontwikkelen van de formules ziet er als volgt uit. Van elke tekst wordt de leesbaarheid, opgevat als begripelijkheid, bepaald; dit is de criteriumwaarde. Voor de operationalisering van het criterium worden onder meer gebruikt: scores op begripend leestoetsen, oordelen van deskundigen, scores op cloze-toetsen (teksten waarbij systematisch woorden zijn weggelaten die door de leerlingen vervolgens ingevuld moeten worden).

Vervolgens worden van de teksten de waarden op de onafhankelijke taalvariabelen (de predictoren) bepaald. Via multiële regressievergelijking wordt daarna vastgesteld welke onafhankelijke variabelen de verschillen in leesbaarheid voorspellen, en welk gewicht de diverse predictoren hebben bij die voorspelling. Het valt buiten het bestek van dit proefschrift om hier nader op in te gaan. Voor een overzicht verwijzen we naar Staphorsius (1994).

Het idee achter de meeste leesbaarheidsformules is veelal het op elkaar afstemmen van de begripende leesvaardigheid van de leerlingen en de leesbaarheid (in de betekenis van begripelijkheid) van de teksten. De scores van de leerlingen op een toets kunnen op basis van de resultaten van het leesbaarheidsonderzoek worden omgezet in een uitspraak die aangeeft welke tekst een leerling wel en niet met begrip kan lezen. Vandaar ook dat

veel formules de leesbaarheid van de teksten omzetten in 'grade equivalent scores' (vergelijkbaar met de didactische leeftijdsequivalenten). De AVI-niveau-indeling is eveneens opgesteld met als doel het op elkaar afstemmen van de moeilijkheidsgraad van teksten en de leesvaardigheid van de leerlingen, met dit verschil dat die indeling bedoeld is voor de *technische* leesvaardigheid. We zullen verderop in dit hoofdstuk hier nader op ingaan.

Welke factoren dragen nu bij aan de leesbaarheid van de tekst? De inhoud, de stijl, het format en algemene organisatiekenmerken zijn daarvoor van belang (Klare, 1984). In de meeste formules zijn geen van deze factoren echter terug te vinden, omdat het moeilijk zo niet onmogelijk was deze factoren objectief te meten. Daarbij speelde nog mee dat men formules wilde maken die eenvoudig toe te passen waren. De meeste ontwikkelaars van formules hebben zich daarom beperkt tot variabelen die terug te brengen zijn tot twee factoren: de semantische en de syntactische factor.

De semantische factor betreft de moeilijkheid of complexiteit van de gebruikte woorden. In de meeste formules is woordfrequentie opgenomen als indicator voor woordmoeilijkheid, ter operationalisering waarvan de woordlengte in aantal letters of lettergrepen of een woordenlijst met bekende veel voorkomende woorden wordt gebruikt.

Voor de syntactische factor, de complexiteit van zinnen, wordt meestal de zinslengte als indicator gebruikt (aantal woorden per zin).

In navolging van de Amerikaanse formule-ontwikkelaars, werden in de jaren zestig ook in Nederland diverse leesbaarheidsformules gemaakt. De zeer populaire Reading Ease formule van Flesch (1948; zie Brouwer, 1963) is daarbij een belangrijk uitgangspunt geweest. Deze formule hebben we heronder weergegeven.

$$RE = 206,835 - 0,846 * WL - 1,015 * ZL \quad (2.1)$$

waarbij :

$WL = 100 * \text{aantal lettergrepen} / \text{aantal woorden}$

$ZL = \text{aantal woorden} / \text{aantal zinnen}$

Douma (1960, zie Brouwer, 1976) paste deze Flesch-formule aan voor de Nederlandse taal door een 10%-correctie toe te passen op de factoren voor woord- en zinslengte. Hij kwam tot deze correctiefactor door vergelijking van de gemiddelde woord- en zinslengte in Engelse teksten en de Nederlandse vertalingen daarvan.

Brouwer (1963) nam voor zijn Leesindex A eveneens de formule van Flesch als uitgangspunt, maar maakte niet zonder meer een vertaling naar het Nederlands. Op basis van de bevindingen van Flesch nam hij de gemiddelde woord- en zinslengte als uitgangspunt voor zijn formule. In afwijking van de meeste leesbaarheidsformules ontwikkelde hij zijn Leesindex A niet met behulp van multiële regressie van een criteriumvariabele op woord- en zinslengte om zo de gewichten van deze twee factoren te bepalen. In zijn onderzoek ging hij uit van een steekproef van 22 boeken uit de Nederlandse lectuur ingedeeld in 12 categorieën, van kinderlectuur voor zes tot acht jaar tot en met studieboeken. Van ieder boek bepaalde hij de gemiddelde woord- en zinslengte. De laagste en hoogste gevonden waarden betroffen 1,33 respectievelijk 1,96 voor het aantal lettergrepen per woord en 6 respectievelijk 27 voor het aantal woorden per zin.

De formule voor de Leesindex A heeft Brouwer afgeleid door uit te gaan van een drietal veronderstellingen:

- 1 de leesbaarheid is het kleinst bij een zo laag mogelijk gemiddeld aantal woorden per zin en gemiddeld aantal lettergrepen per woord; voor dit geval wordt de Leesindex A op 100 gesteld;
- 2 de leesbaarheid is het grootst bij een zo hoog mogelijk gemiddeld aantal woorden per zin en gemiddeld aantal lettergrepen per woord; voor dit geval wordt de Leesindex A op 0 gesteld;
- 3 de invloed van de woordlengte op de leesbaarheid is even groot als de invloed van de zinslengte.

Dit laatste wijkt af van de formule van Flesch, waarin de woordlengte zwaarder weegt dan de zinslengte. Brouwer wijzigde dit op basis van zijn eigen indruk dat de woord- en zinslengte ongeveer een even grote invloed hebben op de moeilijkheid van teksten. Deze drie veronderstellingen heeft hij, gebruikmakend van de gevonden laagste en hoogste gemiddelden voor woord- en zinslengte, omgezet in drie vergelijkingen:

$$\text{ad 1: } a + 127b + 5c = 100$$

$$\text{ad 2: } a + 202b + 30c = 0$$

$$\text{ad 3: } (202 - 127) * b = (30 - 5) * c$$

Daarbij heeft Brouwer, om 'mooie' gewichten te verkrijgen, de waarden voor woord- en zinslengte aangepast. Via simpele substitutie heeft hij Leesindex A verkregen (zie formule 2.2).

$$\text{Leesindex A} = 195 - (2/3) * \text{WL} - 2 * \text{ZL} \quad (2.2)$$

waarbij :

WL = $100 * \text{aantal lettergrepen} / \text{aantal woorden}$

ZL = $\text{aantal woorden} / \text{aantal zinnen}$

In later ontwikkelde Nederlandse leesbaarheidsformules is wel de weg van de multiële regressie gevolgd. Zo hebben Zondervan, Van Steen en Gunneweg (1976) de leesbaarheid van teksten op twee manieren geoperationaliseerd. Op de eerste plaats via leerkrachtoordelen: de leerkrachten moesten bij elke tekst aangeven of deze te makkelijk, geschikt of te moeilijk was voor een leerjaar. Op de tweede plaats via toepassing van de cloze-procedure bij leerlingen van verschillende leerjaren. Zowel de leerkrachtoordelen als de cloze-scores waren de afhankelijke variabelen in hun onderzoek. Als onafhankelijke variabelen werden naast woord- en zinslengte nog 12 andere tekstenmerken gerelateerd aan de leesbaarheid van de tekst. Met behulp van multiële regressie-analyses ontwikkelden ze diverse formules, waaronder aparte formules voor leerjaar 3 tot en met 6 ten behoeve van het gebruik in de klas.

Hun onderzoek leverde overigens gegevens op over de empirische validiteit van de formules van Brouwer en Douma: beide correleerden hoog met de leerkrachtoordelen (.90-.91 resp. .84-.86) en de clozescores van de leerlingen (-.84 respectievelijk -.82). De beide formules bleken in de praktijk te werken voor de ordening van teksten naar moeilijkheidsgraad.

De meest recente leesbaarheidsformule voor de Nederlandse taal is de CLIB (Cito LeesIndex voor het Basis- en speciaal onderwijs, zie Staphorsius, 1994). Het doel van de ontwikkeling van deze index samen met de bijbehorende CLIB-toetsen was hetzelfde als bij de ontwikkeling van de meeste formules: het op één schaal brengen van de begrijpelijkheid van teksten en de begrijpende leesvaardigheid van leerlingen. Voor het ontwikkelen van de CLIB is een onderzoek uitgevoerd naar de leesbaarheid van teksten bij een corpus van 240 non-fictie teksten voor 7 t/m 12 jarigen. De leesbaarheid van de teksten is vastgesteld met behulp van de cloze-procedure. Door middel van stapsgewijze multiële regressie is de leesbaarheidsformule ontwikkeld, waarbij de gemiddelde clozescore (het gemiddeld aantal goed ingevulde woorden) de afhankelijke variabele was. De onafhankelijke variabelen, de predictoren, betroffen, net als in andere leesbaarheidsonderzoeken, verschillende operationalisering van de lexicale en de syntactische complexiteit. Daarbij is binnen de lexicale factor nog een onderscheid gemaakt tussen de woordlengte, de woordfrequentie/woordbekendheid en de diversiteit van de woordenschat. De syntactische complexiteit betrof enerzijds verschillende operationalisering van de zinslengte en anderzijds de variabelen tanglengte (aantal woorden respectievelijk aantal lettergrepen per tang) en tangdichtheid (percentage tangen per zin)¹.

In de uiteindelijke formule (zie figuur 2.3) zijn vier onafhankelijke variabelen opgenomen:

- het aantal letters per woord (ter operationalisering van de factor 'woordlengte');
- het percentage woorden dat uit een bepaalde woordbekendheidslijst afkomstig is (factor 'woordbekendheid');
- het percentage types per token, ofwel de verhouding tussen het aantal *verschillende* woorden en het totaal aantal woorden bij de eerste 250 woorden van de tekst (factor 'diversiteit woordenschat');
- het percentage zinnen per woord ('zinslengte').

$$CLIB = 46 - 6,603 * GWL + 0,474 * FREQ77 - 0,365 * PTYPES + 1\,425 * PZW \quad (2.3)$$

waarbij :

GWL = aantal letters /aantal woorden

FREQ77 = 100 * aantal woorden in woordenlijst / totaal aantal woorden

PTYPES = 100 * aantal verschillende woorden / totaal aantal woorden (van de eerste 250 woorden)

PZW = 100 * aantal zinnen / aantal woorden

De resulterende CLIB (zie formule 2.3) rangordent de teksten naar leesbaarheid. De overeenkomst tussen de Leesindex A en de CLIB is dat in beide de factoren 'woordlengte' en 'zinslengte' zijn opgenomen. De CLIB bevat daarnaast ook de factoren 'woordbekendheid' en 'diversiteit woordenschat': als een hoog percentage bekende woorden of een hoog percentage dezelfde woorden in een tekst wordt gebruikt, zal de tekst gemakkelijker zijn.

¹ Staphorsius beperkte zich tot twee tangconstructies: de persoonsvorm-rest-constructie (deel van de zin tussen persoonsvorm en het laatste daarbij behorende voltooid deelwoord) en de onderwerp-persoonsvorm-constructie (deel van de zin tussen onderwerp en de persoonsvorm).

In de loop der tijd zijn veel leesbaarheidsformules ontwikkeld en worden de formules door velen gebruikt. Er is echter ook veel kritiek op de leesbaarheidsformules en het gebruik daarvan:

- belangrijke variabelen die de leesbaarheid bepalen zijn er meestal niet in opgenomen, zoals bijvoorbeeld de inhoud, de stijl, het gebruikte vocabulaire, de ordening van de tekst, de complexiteit van ideeën, de achtergrondkennis, motivatie en interesse van de lezer;
- de wel opgenomen variabelen zijn niet adequaat; een langere zin is niet altijd moeilijker dan een kortere, de volgorde van woorden in een zin doet er in de meeste formules niet toe;
- toepassing van de verschillende formules levert inconsistente resultaten: volgens de ene formule is tekst A makkelijker dan tekst B, volgens de andere is het net andersom; volgens de ene formule is tekst A geschikt voor groep 4 en volgens de andere voor groep 6.

Waarom zouden we eigenlijk leesbaarheidsformules gebruiken voor het bepalen van de moeilijkheidsgraad van teksten, zeker gezien bovengenoemde kritiek? Er zijn immers ook andere methoden om de leesbaarheid te schatten, bijvoorbeeld door de cloze-procedure toe te passen op teksten.

Klare (1984) wijst in dit kader echter op het verschil tussen beide methoden: de formules *voorspellen* de leesbaarheid van een tekst, de cloze-procedure en andere methoden gericht op tekstbegrip, *meten* de leesbaarheid. Het toepassen van de cloze-procedure vereist dat veel proefpersonen de teksten lezen. Voor het toepassen van de leesbaarheidsformules hoeft dat niet, waardoor dat minder tijd en moeite kost. De cloze-procedure kan wel goed gebruikt worden als criterium voor leesbaarheid bij het ontwikkelen van leesbaarheidsformules, zoals bijvoorbeeld Staphorsius (1994) heeft gedaan.

Een andere mogelijkheid voor het vaststellen van de leesbaarheid van teksten is het gebruiken van beoordelaars. Uit onderzoek blijkt echter dat hun oordelen nogal uiteen kunnen lopen. Klare ziet beoordelaars als complementair ten opzichte van de leesbaarheidsformules. De formules kunnen gebruikt worden als eerste screening. Met beoordelaars kan vervolgens worden voorkomen dat de formules en de uitkomsten ervan misbruikt worden door hen de teksten te laten nalezen op thema, stijl, ordening, enzovoorts.

Een deel van de kritiek op de leesbaarheidsformules is ook te herleiden tot het verkeerde gebruik ervan. De leesbaarheidsformules zijn ontwikkeld voor het *voorspellen* van leesbaar schrijven, en niet voor het *produceren* van leesbaar schrijven (Klare, 1984). Natuurlijk kunnen schrijvers de formules 'voor de gek houden' door bijvoorbeeld woorden of zinnen van een tekst door elkaar te gooien: de uitkomst van de formule verandert niet, maar de leesbaarheid wel. Een veel gehoorde kritiek is het schrijven 'volgens de formule': teksten worden aangepast om een bepaalde leesbaarheidsindex te krijgen of teksten worden geheel geschreven volgens de formule. De leesbaarheidsformules zijn hiervoor niet bedoeld. Ze zijn bedoeld voor het voorspellen van de moeilijkheidsgraad van een tekst en niet voor het schrijven van een tekst. Voor het voorspellen is het voldoende dat de onafhankelijke variabelen een correlatieel verband hebben met leesbaarheid, voor het produceren van teksten moet er echter een causale relatie bestaan. De onafhankelijke variabe-

len in de leesbaarheidsformules zijn geen oorzaken van de verschillen in leesbaarheid, maar zijn kenmerken die in normale teksten variëren met de moeilijkheidsgraad. De ervaring leert dat in moeilijkere teksten, bestemd voor meer ervaren lezers, de zinnen gemiddeld genomen langer zijn, minder frequent voorkomende woorden gebruikt worden, enzovoorts. Van dit covariëren van dergelijke tekstkenmerken en inhoudelijke complexiteit maakt men in leesbaarheidsformules gebruik om de moeilijkheidsgraad van teksten te *voorspellen*.

Afsluitend kunnen we concluderen dat het gebruik van leesbaarheidsformules een geschikt hulpmiddel kan zijn bij het bepalen van de moeilijkheidsgraad van een tekst. Tegelijkertijd moeten we echter geen absolute waarde aan de uitkomsten hechten. Bij de uitkomsten van de formules moet altijd rekening worden gehouden met een betrouwbaarheidsmarge, waardoor we deze uitkomsten alleen als ruwe indicatoren voor de leesbaarheid mogen zien. De leesbaarheidsindex mag ook nooit het enige criterium zijn op basis waarvan boeken en leesstof worden geselecteerd. Tevens is het niet verantwoord om de leesbaarheidsformules als basis te gebruiken voor het schrijven van boeken en teksten. Zoals net is gezegd hebben variabelen zoals woord- en zinslengte geen causale relatie met de leesbaarheid van een tekst. Zeer belangrijk zijn de organisatie van een verhaal, het format, de stijl, de inhoud, enzovoorts.

2.1.2 Leesindex A en de negen AVI-niveaus

We hebben gezien dat de Leesindex A van Brouwer een belangrijk uitgangspunt voor de negen AVI-niveaus geweest is. In bijlage 3 hebben we de oorspronkelijke indeling van de niveaus weergegeven. De verschillen tussen de eerste vier niveaus zijn terug te voeren op specifieke woord- en tekstkenmerken (bijvoorbeeld in niveau 1 alleen eenlettergrepige woorden en slechts één zin per regel) en slechts gedeeltelijk op de Leesindex A. De indeling van de overige vijf niveaus is geheel gebaseerd op de Leesindex A. Hoe is nu deze indeling op basis van de leesindex tot stand gekomen? Een onderzoek van Brouwer uit 1976 ligt daaraan ten grondslag.

In het eerste onderzoek van Brouwer (1963), waarin hij de Leesindex A ontwikkelde, werd de gemakkelijkste categorie gevormd door kinderlectuur tot en met 12 jaar met bijbehorende Leesindex A van 80 tot en met 100. Deze indeling was echter te grof voor het gebruik in de basisschool. Daarom maakte Brouwer (1976) een fijnere verdeling van de leesmoeilijkheid ten behoeve van het *voortgezet leesonderwijs*. Hij bepaalde van 75 leesseries en leesboekjes voor de toenmalige leerjaren 2 tot en met 6 de gemiddelde woord- en zinslengte. In tabel 2.1 hebben we zijn bevindingen weergegeven.

Tabel 2.1: Overzicht gemiddelde zins- en woordlengte van de onderzochte leesboekjes per leerjaar, alsmede de uiterste waarden ervan (uit: Brouwer, 1976)

Klas [groep]	gemiddelde zinslengte* (min-max)	gemiddelde woordlengte* (min-max)
2 (N=12) [4]	7 (5-8)	129 (123-143)
3 (N=11) [5]	9 (7-10)	135 (128-140)
4 (N=15) [6]	10 (7-14)	137 (129-140)
5 (N=19) [7]	10 (8-15)	143 (134-152)
6 (N=18) [8]	11 (9-17)	144 (133-163)

NB: * zinslengte = aantal woorden/aantal zinnen;

woordlengte = 100 * aantal lettergrepen/aantal woorden

Uit deze tabel blijkt duidelijk de grote spreiding in woord- en zinslengte per leerjaar en de overlap tussen de leerjaren. De gegevens waren voor Brouwer aanleiding te veronderstellen dat als ondergrens voor technisch lezen een Leesindex van 70 reëel is. In het betreffende artikel staat niet weergegeven hoe hij hiertoe komt, maar we kunnen zijn redenering wel afleiden. Brouwer veronderstelde namelijk dat in klas 4 (groep 6) het technisch lezen beheerst dient te worden. De hoogste woord- en zinslengte voor klas 4 leidden tot een Leesindex van 72 en deze heeft hij waarschijnlijk afgerond tot 70. Vervolgens verdeelde hij de range van 70 tot 100 in gelijke stukken voor de drie leerjaren (zie ook tabel 1.2 in paragraaf 1.1.1) en bestemde hij boekjes met een Leesindex tussen 90 en 99 voor leerjaar 2 (groep 4), tussen 80 en 89 voor leerjaar 3 (groep 5) en tussen 70 en 79 voor leerjaar 4 (groep 6). Alle boekjes met een Leesindex groter dan 100 waren naar de mening van Brouwer bedoeld voor het *aanvankelijk leesonderwijs* in leerjaar 1 (groep 3). Gezien de wijze waarop de Leesindex A is geconstrueerd en vervolgens de indelingen per leerjaar zijn vastgesteld, kunnen we vraagtekens zetten bij de betrouwbaarheid van de indeling. Zo is sprake van een grote spreiding binnen een leerjaar en een grote overlap tussen de leerjaren (zie tabel 2.1).

Deze indeling naar Leesindex A zien we terug in de indeling van de negen AVI-niveaus. Van den Berg en Te Lintelo (1977b) hebben voor *de eerste vier AVI-niveaus*, bedoeld voor het aanvankelijk lezen (zie paragraaf 1.1.1), een onderscheid gemaakt in de Leesindex tussen de niveaus 1 en 2 enerzijds (110 en hoger) en de niveaus 3 en 4 anderzijds (109-100). Maar voor de eerste vier niveaus speelt de Leesindex A een geringe rol. Het meest belangrijk zijn de specifieke woord- en tekstkenmerken (bijvoorbeeld alleen eenlettergrepige woorden in niveau 1; zie bijlage 3). Wat opvalt in de indeling zijn de uitzonderingen. Bijvoorbeeld het woord 'moeder' in AVI-niveau 1, waarin verder alleen eenlettergrepige woorden zijn toegestaan. Of het woord 'vader' in AVI-niveau 2, waarin voor het overige alleen tweelettergrepige woorden zonder spellingsmoeilijkheden zijn toegestaan.

AVI-NIVEAU 1:

- Tekstkenmerken:**
- korte zinnen, één zin per regel;
 - samengestelde zinnen over twee regels verdeeld, komen voor (voor Leesindex geteld als twee zinnen);
 - hoofdletters kunnen voorkomen;
 - Leesindex A: > 100.

Woordtypen:

- éénlettergrepige woorden:
 - . medeklinker-klinker: zo, pa;
 - . klinker-medeklinker. ik, om;
 - . medeklinker-klinker-medeklinker. naam, roos;
- éénlettergrepige woorden met één medeklinkercombinatie kunnen voorkomen.

AVI-NIVEAU 2:

- Tekstkenmerken:**
- hoofdletters kunnen voorkomen;
 - korte zinnen, één zin per regel;
 - samengestelde zinnen worden nog per zin over twee regels verdeeld (voor de Leesindex geteld als twee zinnen);
 - Leesindex A: > 100.

**Woordtypen:
(uitbreiding)**

- éénlettergrepige woorden:
 - . eindigend op dt, ng of nk: houdt, bang, zink;
 - . twee medeklinkers voor en/of achteraan het woord: stoel, pats, stelt;
 - . drie medeklinkers voor of achteraan het woord: straat, kamp;
- tweelettergrepige of samengestelde woorden zonder spellingsmoeilijkheden: voetbal, mamma;
- verkleinwoorden: huisje.

AVI-NIVEAU 3:

- Tekstkenmerken:**
- korte zinnen kunnen betekenisvol zijn afgebroken en doorlopen op de volgende regel. De nieuwe zin begint nog wel op een nieuwe regel;
 - Leesindex A > 100.

**Woordtypen:
(uitbreiding)**

- alle typen éénlettergrepige woorden;
- alle typen tweelettergrepige woorden;
- drie- en meerlettergrepige woorden zonder spellingsmoeilijkheden.

AVI-NIVEAU 4:

- Tekstkenmerken:**
- zinnen kunnen, betekenisvol afgebroken, doorlopen op de volgende regel;
 - een gesproken zin begint nog steeds op een nieuwe regel;
 - Leesindex A > 100.

**Woordtypen:
(uitbreiding)**

- alle typen drie- en meerlettergrepige woorden;
- leenwoorden zonder lastige, afwijkende teken-klankcombinaties: portemonnee, diskette.

AVI-NIVEAU 5:

- Tekstkenmerken:**
- Leesindex A: 99-94;
 - gemiddelde zinslengte: 7 woorden (6-8 woorden);
 - gemiddelde woordlengte: 127 lettergrepen (123-130).

AVI-NIVEAU 6:

- Tekstkenmerken:**
- Leesindex: A 93-89;
 - gemiddelde zinslengte: 8 woorden (7-9 woorden);
 - gemiddelde woordlengte: 132 lettergrepen (129-135).

AVI-NIVEAU 7:

- Tekstkenmerken:**
- Leesindex A: 88-84;
 - gemiddelde zinslengte: 9 woorden (8-10);
 - gemiddelde woordlengte: 136 lettergrepen (134-139).

AVI-NIVEAU 8:

- Tekstkenmerken:**
- Leesindex A: 83-79;
 - gemiddelde zinslengte: 10 woorden (9-11);
 - gemiddelde woordlengte: 141 lettergrepen (138-144).

AVI-NIVEAU 9:

- Tekstkenmerken:**
- Leesindex A: 78-74;
 - gemiddelde zinslengte: 11 woorden (10-12);
 - gemiddelde woordlengte: 146 lettergrepen (143-148).

Figuur 2.1: Nieuwe indeling van de negen AVI-niveaus

Deze uitzonderingen zijn het gevolg van het feit dat de opbouw van de methode Veilig Leren Lezen als uitgangspunt is genomen, hetgeen overigens goed te verdedigen was gezien het marktaandeel van de methode destijds (circa 85%).

Het onderscheid in Leesindex A voor de vier laagste AVI-niveaus hebben we voor de nieuwe AVI-toets losgelaten, mede omdat de andere kenmerken het meest bepalend zijn; voor alle vier niveaus geldt een Leesindex van 100 of meer. Woord- en zinskenmerken bepalen of een tekst tot AVI-niveau 1, 2, 3 of 4 behoort. Getracht is de betreffende criteria zodanig bij te stellen dat deze zo min mogelijk methode-afhankelijk zijn. Uitzonderingen, zoals het boven genoemde 'moeder' in AVI-niveau 1, zijn verwijderd, alsmede het criterium betreffende het gebruik van hoofdletters. In figuur 2.1 (alsmede in bijlage 1) zijn de herziene criteria opgenomen. Opgemerkt moet worden dat deze criteria voor de teksten van de toets strikt toegepast zijn. Voor het indelen van boeken naar AVI-niveau zullen de criteria wat soepeler worden gehanteerd. Immers, wat is een enkel moeilijk woord op een totaal van 1000 woorden?

De opbouw van de AVI-niveaus 1 tot en met 4 naar type woorden loopt grotendeels parallel met de leergang van de methoden voor aanvankelijk lezen: AVI-niveau 1 betreft teksten met uitsluitend eenvoudige eenlettergrepige woorden en pas in AVI-niveau 4 komen teksten voor met drie- of meerlettergrepige woorden (Ellerman & Spaaij, 1986). We betwijfelen echter of de technische leesvaardigheid van de gemiddelde leerling in de periode van het aanvankelijk leesonderwijs (groep 3) zodanig ontwikkeld is, dat hij of zij (de woorden van) de eerste vier AVI-niveaus automatisch en snel kan herkennen. De technische leesvaardigheid van de gemiddelde leerling zal aan het einde van groep 3 naar alle waarschijnlijkheid niet gekenmerkt kunnen worden door 'AVI-niveau 3 of 4', alhoewel de eerste vier niveaus oorspronkelijk wel bedoeld zijn voor het aanvankelijk leesniveau. Dit wordt duidelijk wanneer we kijken naar de in paragraaf 1.2.4 besproken drie fasen in de ontwikkeling van woordherkenning: de nauwkeurigheidsfase, de automatische fase en de snelheidsfase (Ehri & Wilce, 1979, 1983).

In de eerste fase besteden de kinderen veel aandacht aan het verwerken van letters en de letter-klank-koppeling, in de tweede fase herkennen ze de woorden zonder aandacht voor de afzonderlijke letters en de letter-klank-koppeling en in de derde fase neemt de snelheid van herkenning toe. Wanneer we nu de tijd- en foutennormen die bij de AVI-toetskaarten worden gesteld (zie paragraaf 1.1.2) naast deze drie fasen leggen, dan constateren we het volgende. Om een bepaald AVI-niveau te beheersen moet een kind de tekst van de bijbehorende toetskaart, en dus de woorden waaruit de tekst is opgebouwd, in ieder geval 'automatisch' kunnen lezen (de tweede onderscheiden fase). Immers te veel aandacht voor de letters en de letter-klank-koppeling (de eerste fase) leidt al gauw tot een prestatie die als onvoldoende wordt beoordeeld:

- het letter-voor-letter spellen van een woord zonder het woord in één keer goed te lezen, wordt fout geteld; als dit veel voorkomt voldoet de leerling niet aan de foutennorm;
- het veelvuldig spellen van een woord, voor het alsnog in één keer te lezen, zal de leestijd doen toenemen, waardoor de leerling niet aan de tijdsnorm voldoet.

Het gebruik van de tijdsnorm bij de AVI-toets geeft bovendien aan dat het decoderen van de betreffende teksten redelijk snel moet verlopen, ofwel dat voor deze teksten de

snelheidsfase bereikt dient te zijn. Aangezien voor het bereiken van een snelle decodeervaardigheid vooral veel oefenen van belang is (Näslund & Samuels, 1992; Perfetti, 1985; Stanovich, 1991), vragen we ons af of de gemiddelde leerling aan het eind van groep 3 woorden behorend tot AVI-niveau 4 al zodanig beheerst dat de snelheidsfase is bereikt. Immers in de meeste gevallen hebben kinderen dergelijke woorden pas aan het einde van groep 4 geleerd.

Al zetten we vraagtekens bij het bereiken van de eerste vier niveaus na afloop van het aanvankelijk leesonderwijs in groep 3, de opbouw van deze niveaus analoog aan de opbouw van de moeilijkheidsgraad van woorden in de methoden voor aanvankelijk leesonderwijs, lijkt ons wel een goed uitgangspunt. In paragraaf 1.3.2 is immers al ter sprake gekomen dat een leerling niet alle woorden tegelijkertijd automatisch en snel zal lezen, maar dat dit geleidelijk zal gebeuren. De gemakkelijkste woorden en de woorden die het eerst geoefend worden, zullen het eerst automatisch en vervolgens snel gedecodeerd worden.

De *AVI-niveaus 5 tot en met 9* zijn door Van den Berg en Te Lintelo (1977b) bedoeld voor het voortgezet technisch leesonderwijs. In een eerdere publikatie (Van den Berg, 1972) werden deze niveaus van elkaar onderscheiden op basis van het aandeel twee- en meerlettergrepige woorden in de teksten, de lengte van de zinnen, de moeilijkheid van de inhoud en de typografie. Ter illustratie:

"Niveau 6:

Boekjes met over het algemeen:

- groter aantal twee- en meerlettergrepige woorden dan in niveau 5;
- korte, doch ook enkele andere langere zinnen;
- moeilijkere inhoud dan de boekjes van niveau 5;
- geen verdeling in lettergrepen; in de meeste boekjes is de tekst ingedeeld in alinea's: vrij klein lettertype." (Van den Berg, 1972, bijlage III-1).

Na twee jaar praktijkervaring met deze indeling wilde men de moeilijkheidsgraad van teksten voor deze niveaus op meer wetenschappelijke manier bepalen. Daarvoor werd steun gezocht bij Brouwer (1976). Brouwer maakte alleen een onderscheid naar leerjaar; de vijf AVI-niveaus zijn daar nog een verdere opdeling van:

AVI-5: Leesindex 99-94;

AVI-6: Leesindex 93-89;

AVI-7: Leesindex 88-84;

AVI-8: Leesindex 83-79;

AVI-9: Leesindex 78-74.

De AVI-niveaus 5 en 6 zijn een verdere opdeling van wat Brouwer (1976) geschikt achtte voor het toenmalige leerjaar 2; de niveaus 7 en 8 zijn dat voor leerjaar 3 en niveau 9 betreft een deel van de range geschikt voor leerjaar 4.

Bij de indeling van de Leesindex A naar leerjaar hebben we reeds onze twijfel geuit over de betrouwbaarheid daarvan. Hetzelfde geldt uiteraard ook voor de indeling van de vijf hoogste AVI-niveaus, die daarop gebaseerd is. De verschillen in de leesindices tussen de aangrenzende AVI-niveaus zijn nogal klein, zeker gezien de grote spreiding van Leesindex A binnen een leerjaar en de overlap tussen de leerjaren, zoals in tabel 2.1 is weer-

gegeven. Dit doet vermoeden dat de moeilijkheidsgraad van teksten zoals bedoeld niet overeen hoeft te komen met de daadwerkelijk ervaren moeilijkheid.

Desondanks hebben we voor de nieuwe AVI-toetskaarten de indeling naar Leesindex A gehandhaafd. We zullen in ons onderzoek nagaan of de ordening van de toetskaarten naar AVI-niveau overeenkomt met de moeilijkheidsgraad zoals blijkt uit de prestaties van de leerlingen. Gezien de grote verspreiding van de bestaande indeling van de AVI-niveaus - veel uitgeverijen en (school)bibliotheken hebben hun boeken ingedeeld naar AVI-niveau - was het niet wenselijk de indeling ineens te verlaten of ingrijpend te wijzigen. Daarvoor zou een uitgebreid onderzoek nodig zijn.

Tot slot willen we nog een laatste probleem vermelden ten aanzien van het gebruik van de Leesindex A voor het ordenen van teksten voor technisch lezen. Leesbaarheidsformules zijn bedoeld om de leesbaarheid of beter gezegd de begrijpelijkheid van teksten vast te stellen. Brouwer (1976) heeft zijn Leesindex A in verband gebracht met het technisch lezen en voor AVI is dat systeem overgenomen. Men kan zich afvragen of het wel correct is om leesbaarheidsformules, die bedoeld zijn voor het vaststellen van de begrijpelijkheid, te gebruiken om teksten te selecteren voor het technisch lezen. Zou dan niet veel eerder iets als 'technische moeilijkheidsgraad' van teksten gemeten moeten worden? Op de eerste plaats vragen we ons af of de ordening van *teksten* naar moeilijkheidsgraad wel verschilt voor technisch lezen en begrijpend lezen. Er is overigens nauwelijks onderzoek gedaan naar een ordening van teksten naar de moeilijkheidsgraad voor het technisch lezen. Brouwer spreekt ook niet over de technische moeilijkheidsgraad van teksten, maar over de geschiktheid van teksten voor het technisch leesonderwijs.

Op de tweede plaats willen we opmerken dat de Leesindex A, net als vele andere leesbaarheidsformules, juist de technische aspecten van een tekst gebruikt om de begrijpelijkheid te bepalen, te weten de woord- en zinslengte. Van de woordlengte is bekend dat deze samenhangt met de snelheid van woordherkenning: korte woorden worden eerder automatisch en snel herkent dan lange woorden (zie paragraaf 1.3.2.).

Samenvattend kunnen we vaststellen dat voor de nieuwe AVI-toetskaarten de oorspronkelijke criteria zoveel mogelijk gehandhaafd zijn. Voor de eerste vier niveaus zijn de methode-afhankelijke criteria verwijderd, voor de overige vijf niveaus is de indeling hetzelfde gebleven. We hebben vastgesteld dat de indeling voor de eerste vier niveaus plausibel lijkt, alhoewel we niet verwachten dat deze niveaus door de gemiddelde leerling al in groep 3 beheerst zullen worden. Bij de indeling van de vijf hoogste niveaus hebben we twijfels bij het onderscheidend vermogen van de Leesindex. In hoofdstuk 3 onderzoeken we of de veronderstelde moeilijkheidsgraad van de toetskaarten overeenkomt met de waargenomen moeilijkheidsgraad.

2.2 De constructie van de toetskaarten

De constructie van de toetskaarten betrof het schrijven van nieuwe teksten op een zodanige manier dat:

- de inhoud van teksten beter aansluit bij de hedendaagse beleavingswereld van kinderen;

- de kaarten van oplopende moeilijkheidsgraad zijn, en
- de kaarten representatief zijn voor de teksten van boekjes van dat niveau.

In de vorige paragraaf hebben we reeds de licht gewijzigde indeling naar negen niveaus beschreven. Voor het overige is de opzet van de nieuwe toets gelijk aan die van de oude. De leestijd en het aantal fouten bepalen de indeling naar AVI-niveaus van de leerlingen. Ook de criteria voor wat fout en niet fout wordt gerekend zijn gehandhaafd (zie paragraaf 1.1.2 en bijlage 4).

Teneinde een zo betrouwbaar en valide mogelijk instrument te maken, zijn tijdens de constructiefase de volgende stappen gezet:

- 1 schrijven van teksten;
- 2 beoordelen van teksten door leesdeskundigen;
- 3 een eerste try-out op 1 school;
- 4 beoordeling van dezelfde teksten door leerkrachten;
- 5 aanpassingen op basis van de ervaringen;
- 6 een tweede try-out op 7 scholen;
- 7 vaststellen van de uiteindelijke teksten.

2.2.1 Nieuwe teksten

De oorspronkelijke AVI-toetskaarten zijn in de jaren zeventig geschreven. Het is dan ook niet verwonderlijk dat de inhoud van de teksten niet meer aansluit bij het huidige tijdsbeeld. Bij het schrijven van de nieuwe teksten hebben de drie auteurs getracht zoveel mogelijk aan te sluiten bij de beleavingswereld van kinderen van nu. De gekozen onderwerpen zijn actueel, bijvoorbeeld ruimtevaart, milieu en televisie. Daarnaast is getracht om teksten te schrijven met een korte verhaallijn en met informatieve waarde voor kinderen uit groep 3 tot en met 6 (Van Laarhoven, 1993). De drie auteurs hebben uiteraard geprobeerd om teksten te schrijven van oplopende moeilijkheidsgraad analoog aan de indeling in negen AVI-niveaus. Ze hebben echter in eerste instantie de teksten niet strikt 'met de criteria in de hand' geschreven. Door de onderzoekers is de eerste versie van iedere tekst gescreend op afwijkingen van de criteria behorend bij het betreffende AVI-niveau. De auteurs hebben de teksten vervolgens aangepast.

De 21 resulterende teksten zijn daarna ter beoordeling voorgelegd aan een aantal leesdeskundigen. De vraag aan deze deskundigen was de teksten kritisch te beoordelen naar verhaallijn, schrijfstijl, begrijpelijkheid en geschiktheid voor kinderen. Ze konden eventueel suggesties te doen voor verbetering, waarbij ze rekening dienden te houden met de opbouw naar AVI-niveaus en met het feit dat de teksten bedoeld zijn voor *toetskaarten*; dat wil zeggen dat het korte, afgeronde teksten moesten zijn die aan alle criteria van een bepaald AVI-niveau dienden te voldoen. De teksten zijn eveneens voorgelegd aan een bureauredactrice van een uitgeverij, die dagelijks betrokken is bij het schrijven, corrigeren en uitgeven van kinderboeken. Zij heeft uiteindelijk de teksten geselecteerd en herschreven, mede gebruikmakend van de opmerkingen en suggesties van de leesdeskundigen. Het resultaat was dat 19 teksten overbleven, drie teksten voor niveau 8 en twee voor de overige niveaus.

2.2.2 De eerste try-out

In de periode van 4 tot en met 8 november 1991 heeft de eerste try-out plaatsgevonden op een basisschool. Vierentachtig leerlingen (34 uit groep 4-5, 28 uit groep 6 en 22 uit groep 7) hebben enkele van de nieuwe kaarten gelezen. Deze try-out was vooral bedoeld om na te gaan of de teksten geschikt waren voor de leerlingen wat betreft woordgebruik en abstractie. De afname geschiedde door drie medewerkers van het KPC, die nauw betrokken waren bij de ontwikkeling van de nieuwe toetskaarten.

De werkwijze was als volgt. In juni 1991 was door leerkrachten van de school voor alle leerlingen het AVI-niveau vastgesteld. Dit niveau diende als uitgangspunt voor de afname van de toetskaarten. Allereerst werd een oude AVI-toetskaart afgenomen, en wel de A-versie van de kaart behorend bij het niveau zoals dat in juni was vastgesteld. De leestijd en het aantal fouten op deze kaart werden genoteerd, zodat nagegaan kon worden in hoeverre de leerling nog op dat AVI-niveau presteerde. Vervolgens werden de A- en B-versie van de nieuwe kaart van hetzelfde niveau aangeboden. Ook hiervan werden de tijd en het aantal fouten genoteerd. Tevens werd nagegaan wat de leerling van de teksten vond en in hoeverre hij of zij de teksten begreep. Daarna werden nog twee kaarten aangeboden, van of een niveau hoger of een niveau lager, in de A- of de B-versie. De laatste van deze twee kaarten was facultatief en werd alleen afgenomen wanneer de tijd en de fitheid van de leerling het toelieten. Welke kaarten een leerling te lezen kreeg was reeds van te voren bepaald, waarbij er voor gezorgd werd dat alle kaarten door ongeveer evenveel leerlingen werden gelezen.

De eerste try-out wees uit dat de teksten nog enige aanpassingen behoefden. De leerlingen bleken bepaalde woorden en zinsconstructies niet te kennen of te begrijpen en enkele teksten waren te abstract. Vergelijking van gemiddelde leestijden en fouten leverden ook aanwijzingen voor verandering op.

2.2.3 Beoordeling door leerkrachten

Om een indruk te krijgen van de moeilijkheidsgraad van de 19 teksten is tevens het oordeel gevraagd van 25 leerkrachten van de groepen 3 tot en met 6, afkomstig van scholen die al jaren met de AVI-toetskaarten werkten. Deze leerkrachten waren goed op de hoogte van de oude AVI-toetskaarten.

Alle 19 teksten zijn, afgedrukt in dezelfde lay-out (lettergrootte, gebruik van hoofdletters), aangeboden aan de leerkrachten met de vraag om van elke kaart het AVI-niveau aan te geven. De leerkrachten is duidelijk gemaakt dat het niet de bedoeling was om na te gaan hoe goed zij de niveaus konden inschatten, maar hoe goed de kaarten volgens hen pasten bij een niveau. De leerkrachten is wel verteld dat er van ieder niveau minimaal twee teksten waren opgenomen, maar tevens is aangegeven dat ze er niet naar moesten streven om per niveau twee teksten aan te wijzen. Het zou mogelijk zijn dat volgens hen geen enkele tekst AVI-niveau 9 heeft en vier teksten niveau 8.

In tabel 2.2 hebben we het gemiddelde en de standaardafwijking van de leerkrachtoordelen weergegeven. Door de leerkrachtoordelen te vergelijken met de niveaus waarvoor de teksten bedoeld zijn, krijgen we een eerste indruk van de moeilijkheidsgraad van de teksten.

Tabel 2.2: AVI-niveaus toegekend door de leerkrachten (N=25)

kaart:	gemiddelde	standaardafwijking	minimum	maximum
1A	1,0	0,2	1	2
1B	1,1	0,3	1	2
2A	2,2	0,4	2	3
2B	2,0	0,2	2	3
3A	3,0	0,4	2	4
3B	3,2	0,5	2	5
4A	4,4	0,6	4	6
4B	4,8	0,9	3	7
5A	5,4	0,6	4	6
5B	5,2	0,9	4	7
6A	5,0	1,0	3	6
6B	5,5	0,9	4	8
7A	7,6	0,8	6	9
7B	7,7	0,8	7	10
8A	7,9	0,8	6	9
8B	7,7	0,8	6	9
8C	7,2	0,7	6	8
9A	8,9	0,5	7	10
9B	8,9	0,4	7	9

De meeste kaarten werden door de leerkrachten gemiddeld op het niveau geschat waarvoor ze bedoeld zijn. Kaart 6A en 6B werden gemiddeld echter op niveau 5 geschat, terwijl kaart 7A en 7B gemiddeld op niveau 8 werden geplaatst. Kaart 8C daarentegen werd door de leerkrachten als makkelijker gezien: niveau 7. Opgemerkt moet worden dat één leerkracht, tegen de instructies in, aan twee kaarten 'hoger dan niveau 9' heeft toegekend.

We hebben Cohens Kappa gebruikt als maat voor de overeenstemming tussen het niveau dat door elke leerkracht is toegekend en het niveau waarvoor de tekst bedoeld is (tabel 2.3). Cohens Kappa geeft de proportie overeenstemmende oordelen weer, gecorrigeerd voor de kans om het goed te raden.

Tabel 2.3: Cohens Kappa voor de overeenstemming van het oordeel van de leerkracht met het niveau zoals bedoeld

Kappa	leerkrachten	
	N	%
0,17	1	4,0
0,23	1	4,0
0,34	1	4,0
0,41	3	12,0
0,47	1	4,0
0,53	5	20,0
0,59	2	8,0
0,64	6	24,0
0,76	1	4,0
0,82	3	12,0
0,88	1	4,0

Het oordeel van twee leerkrachten kwam nauwelijks overeen met de indeling zoals bedoeld (Cohens kappa 0,17 en 0,23); 18 leerkrachten oordeelden redelijk in overeenstemming met de bedoelde indeling (kappa groter dan 0,50, waarvan in vijf gevallen een kappa groter dan 0,75).

Op basis van de resultaten van de try-out bij de leerlingen en de beoordeling door de leerkrachten werden de kaarten nogmaals aangepast door de bureauredactrice, in samenwerking met het onderzoeksteam. Tevens kon een keuze gemaakt worden uit de drie teksten voor niveau 8. Eén van de teksten (8A) was duidelijk makkelijker dan de andere twee; de leerlingen maakten nauwelijks fouten bij het lezen. Deze tekst is dan ook verwijderd. Overigens zou naar de mening van de leerkrachten kaart 8C de gemakkelijkste van de drie kaarten van niveau 8 zijn. Uiteindelijk resulteerden 18 teksten die in de tweede try-out werden uitgetoetst.

2.2.4 De tweede try-out

De tweede try-out vond in januari 1992 plaats op zeven basisscholen en betrof in totaal 601 leerlingen van groep 3 tot en met 7 (zie tabel 2.4). De afname van de kaarten gebeurde in principe door de school zelf. Wie binnen de school de toets zou afnemen, werd aan de leerkrachten zelf overgelaten. Belangrijk was echter dat deze personen ervaring hadden met het afnemen van de oude AVI-toetskaarten. Op een aantal scholen is bijgesprongen door medewerkers van het KPC. Ten behoeve van de afnames is een instructie

geschreven die op de betreffende scholen is toegelicht. In bijlage 4 staat de verkorte versie van deze instructie weergegeven, die tijdens de toetsafnames gebruikt werd.

Tabel 2.4: Aantal leerlingen per school en per groep

school:	groep:					totaal
	3	4	5	6	7	
1	36	32	28	31	8	135
2	24	17	17			58
3	4	8	3	3	2	20
4	3	20	20	16	3	62
5		20	2	14	4	40
6	8	23	23	24	19	97
7	52	58	31	20	28	189
totaal	127	178	124	108	64	601

De werkwijze was als volgt. Eerst is van iedere leerling het AVI-niveau bepaald met behulp van de oude versie A-kaarten. Wanneer de school na 15 november 1991 nog de AVI-niveaus had bepaald, konden deze gegevens als uitgangspunt worden genomen. De tijd en het aantal fouten op de oude AVI-kaart werden genoteerd. Vervolgens kreeg iedere leerling in principe zes nieuwe toetskaarten voorgelegd, drie van de A-versie en drie van de B-versie. De leerlingen kregen te lezen: de kaarten van één niveau lager dan hun AVI-niveau, de kaarten van hun AVI-niveau en de kaarten van één niveau hoger. Een uitzondering werd gevormd door de leerlingen die op AVI-niveau 1 of 9 lezen; deze leerlingen kregen slechts twee kaarten van beide versies te lezen.

De A- en de B-versie van de kaarten zijn bij alle leerlingen afgenomen, waarbij de tijd tussen de twee afnamen klein is gehouden, maar minimaal één dag bedroeg. Op drie scholen kregen de leerlingen versie A als eerste en op de andere vier scholen versie B. In de praktijk hebben niet alle leerlingen beide versies gelezen, doordat ze op één van de toetsdagen afwezig waren. In andere gevallen hebben leerlingen om uiteenlopende redenen niet altijd drie kaarten per versie gelezen, bijvoorbeeld omdat de leerling vermoeid raakte of omdat de afname verstoord werd.

Tijdens de afname van de toetskaarten zijn alle gemaakte fouten op woordniveau en de leestijd per kaart genoteerd. Hiervoor zijn aparte scoreformulieren gemaakt, die optisch gelezen konden worden (zie bijlage 5 voor een voorbeeld). Tijdens de afname werden op dit formulier de fout gelezen woorden aangestreept, waarbij tevens het type fout werd aangegeven. Daarvoor zijn dezelfde zes categorieën gehanteerd als bij de oude AVI-toets (zie paragraaf 1.1.2), namelijk een woord fout gelezen, spellend gelezen, overgeslagen, ongewisseld, tussengevoegd en voorgezegd (zie ook bijlage 4). Voor een uitgebreide

beschrijving van de afnameprocedure, de scoring van de fouten en het gebruik van de scoreformulieren verwijzen we naar Visser, Van Laarhoven en Ter Beek (1996). De leerlingen kregen ruim de tijd voor het lezen van de toetskaarten. Wel werd in principe gestopt met het lezen van een kaart wanneer dit langer dan 5 minuten duurde.

Nadat de leerling de gehele kaart had gelezen, werden op het formulier de leestijd en het totaal aantal gemaakte fouten genoteerd. Voor het optellen van de fouten mocht men een woord slechts één keer fout tellen, ook al maakte een leerling meerdere fouten in hetzelfde woord.

In bijlage 6 hebben we tabellen opgenomen met de descriptieve statistieken voor de leestijd en het aantal fouten per toetskaart. Daarnaast is voor de leestijd nog het gemiddeld aantal gelezen woorden per minuut weergegeven en voor het aantal fouten het percentage fout gelezen woorden. Uit deze gegevens blijkt dat de gemiddelde leessnelheid (het aantal gelezen woorden per minuut) toeneemt met het niveau van de toetskaarten en dus met de vaardigheid van de leerlingen. De gemiddelde leessnelheid is echter niet altijd gelijk voor twee kaarten voor eenzelfde niveau. Het gemiddelde percentage fouten neemt af van de kaarten van niveau 1 tot die van niveau 4; vanaf de kaarten voor niveau 5 is het gemiddelde percentage fouten stabiel (tussen 1,5% en 2,0%). Het percentage fouten op kaarten van eenzelfde niveau is redelijk gelijk.

We zijn nagegaan of de leestijd en het aantal fouten verschilden afhankelijk van het AVI-niveau waarop de leerling volgens de oude toetskaart leest. Dit geeft een indicatie van de moeilijkheidsgraad van de kaarten. We mogen namelijk verwachten dat leerlingen met een hoger AVI-niveau beter zullen presteren bij het lezen van een bepaalde toetskaart dan leerlingen met een lager AVI-niveau. In de tabellen 2.6 tot en met 2.9 hebben we voor versie A respectievelijk versie B de gemiddelde leestijd en het gemiddelde aantal fouten per toetskaart weergegeven, uitgesplitst naar AVI-niveau.

We verwachtten voor de leestijd een lineaire trend: hoe hoger het AVI-niveau van een leerling is, des te minder leestijd een leerling nodig heeft voor een bepaalde toetskaart. Deze verwachting komt over het algemeen uit. Trend-analyse geeft aan dat voor alle toetskaarten minimaal sprake is van de verwachte lineaire trend. Voor sommige toetskaarten geldt dat de lineaire component niet voldoende is ter verklaring van de verschillen in leestijd: er is geen gelijkmatige afname van de gemiddelde leestijd bij een toename van het AVI-niveau van de leerlingen. Zo zien we bijvoorbeeld in tabel 2.6 dat leerlingen die AVI-niveau 1 nog niet beheersen gemiddeld 214 seconden doen over kaart 1A, terwijl leerlingen die lezen op AVI-niveau 1 daar gemiddeld 117 seconden voor nodig hebben; een verschil van 97 seconden. Het verschil in leestijd voor kaart 1A tussen leerlingen van AVI-niveau 1 en 2 bedraagt 25 seconden, en tussen niveau 2 en 3 15 seconden. Ook voor kaart 2A zien we in de tabel een groot verschil in leestijd tussen leerlingen die niveau 1 nog niet beheersen en de leerlingen die dat wel doen (105 seconden), terwijl de verschillen in leestijd tussen de overige niveaus kleiner zijn. Zoals blijkt uit tabel 2.8 is ook voor de kaarten 1B en 2B sprake van een veel groter verschil in leestijd tussen leerlingen die AVI-niveau 1 nog niet beheersen en zij die dat wel doen.

Tabel 2.6: Gemiddelde leestijd in seconden per toetskaart van versie A, uitgesplitst naar het AVI-niveau van de leerlingen volgens de oude kaarten (N tussen haakjes)

AVI-niveau vastge- steld met de oude kaarten	kaartnummer								
	1A	2A	3A	4A	5A	6A	7A	8A	9A
lager dan 1	213,8 (57)	283,8 (38)							
1	117,2 (27)	179,2 (26)	223,0 (5)						
2	91,8 (34)	121,8 (43)	156,3 (33)	133,3 (9)					
3	76,7 (15)	99,9 (33)	138,4 (37)	148,8 (21)	124,0 (4)				
4		93,1 (12)	104,1 (40)	121,5 (42)	118,5 (30)	83,0 (1)			
5			85,1 (8)	95,8 (28)	102,7 (26)	90,9 (20)			
6				87,9 (8)	96,4 (38)	86,3 (41)	134,4 (35)	129,8 (5)	
7					86,7 (6)	77,9 (49)	116,8 (56)	123,2 (50)	125,7 (6)
8						81,9 (16)	107,4 (67)	106,6 (73)	120,5 (56)
9							94,0 (21)	87,7 (137)	101,7 (137)
hoger dan 9								73,6 (22)	88,4 (22)

Tabel 2.7: Gemiddeld aantal fouten per toetskaart van versie A, uitgesplitst naar AVI-niveau van de leerlingen volgens de oude kaarten (N tussen haakjes)

AVI-niveau vastge- steld met de oude kaarten	kaartnummer								
	1A	2A	3A	4A	5A	6A	7A	8A	9A
lager dan 1	14,2 (57)	15,1 (38)							
1	4,6 (27)	7,8 (26)	12,8 (5)						
2	1,8 (34)	2,8 (43)	6,9 (33)	6,3 (9)					
3	1,5 (15)	2,4 (33)	4,7 (37)	7,2 (21)	2,8 (4)				
4		1,8 (12)	2,9 (40)	4,3 (42)	2,1 (30)	4,0 (1)			
5			3,9 (8)	3,1 (28)	2,2 (26)	2,6 (20)			
6				2,3 (8)	2,0 (38)	2,1 (41)	4,6 (35)	4,2 (5)	
7					0,7 (6)	1,6 (49)	3,4 (56)	4,8 (50)	2,0 (6)
8						1,9 (16)	3,6 (67)	4,7 (73)	4,3 (56)
9							3,4 (21)	2,8 (137)	2,8 (137)
hoger dan 9								3,4 (22)	3,3 (22)

Tabel 2.8: Gemiddelde leestijd in seconden per toetskaart van versie B, uitgesplitst naar AVI-niveau van de leerlingen volgens de oude kaarten (N tussen haakjes)

AVI-niveau vastge- steld met de oude kaarten	kaartnummer:								
	1B	2B	3B	4B	5B	6B	7B	8B	9B
lager dan 1	235,1 (56)	289,9 (35)							
1	129,8 (27)	175,9 (25)	225,3 (4)						
2	101,5 (32)	119,4 (42)	166,8 (33)	118,6 (9)					
3	83,9 (16)	95,7 (32)	146,4 (36)	138,9 (20)	133,5 (4)				
4		87,2 (12)	109,8 (40)	112,4 (42)	116,7 (30)	115,0 (1)			
5			92,4 (8)	93,4 (27)	112,3 (28)	104,3 (18)			
6				85,1 (8)	93,0 (35)	87,2 (43)	130,8 (35)	102,4 (5)	
7					84,0 (6)	80,2 (49)	114,3 (55)	102,1 (49)	123,8 (6)
8						69,9 (16)	100,4 (68)	89,8 (73)	124,1 (56)
9							85,7 (20)	74,1 (136)	100,5 (136)
hoger dan 9								61,8 (22)	84,8 (22)

Tabel 2.9: Gemiddeld aantal fouten per toetskaart van versie B, uitgesplitst naar AVI-niveau van de leerlingen volgens de oude kaarten (N tussen haakjes)

AVI-niveau vastge- steld met de oude kaarten	kaartnummer:								
	1B	2B	3B	4B	5B	6B	7B	8B	9B
lager dan 1	13,9 (56)	14,3 (35)							
1	6,0 (27)	7,2 (25)	12,3 (4)						
2	2,3 (32)	1,9 (42)	6,7 (33)	4,2 (9)					
3	1,4 (16)	1,4 (32)	4,8 (36)	5,0 (20)	3,8 (4)				
4		2,3 (12)	3,9 (40)	3,5 (42)	4,2 (30)	4,0 (1)			
5			6,4 (8)	3,1 (27)	4,5 (28)	2,4 (18)			
6				2,3 (8)	3,2 (35)	2,3 (43)	4,9 (35)	5,4 (5)	
7					2,0 (6)	1,8 (49)	3,3 (55)	3,9 (49)	2,0 (6)
8						2,3 (16)	3,6 (68)	2,9 (73)	3,8 (56)
9							4,3 (20)	2,5 (136)	3,1 (136)
hoger dan 9								3,2 (22)	3,7 (22)

Voor de kaarten 4A, 4B en 9B geldt eveneens dat de lineaire component geen voldoende verklaring geeft voor de afname van de leestijden, waarbij sprake is van steeds kleiner wordende verschillen in leestijd naarmate het AVI-niveau van de leerlingen hoger is.

Ten aanzien van het aantal fouten verwachtten we dat er voor de kaarten van de lagere AVI-niveaus sprake is van een lineaire trend. Deze trend verwachtten we niet voor de kaarten van de hogere AVI-niveaus, aangezien de nauwkeurigheid waarmee woorden gelezen worden al snel niet meer verschilt tussen leerlingen (zie hoofdstuk 1). De resultaten van de trendanalyses bevestigen deze verwachtingen. Voor de kaarten 1A, 1B, 2A, 2B vonden we, naast een lineaire trend, nog een andere, die vooral veroorzaakt wordt door het grote aantal fouten gemaakt door de leerlingen die AVI-niveau 1 nog niet beheersen. Zo maken leerlingen die AVI-niveau 1 nog niet beheersen gemiddeld 14 fouten op kaart 1A en 15 fouten op kaart 2A (zie tabel 2.7), terwijl leerlingen die volgens de oude toetskaarten op niveau 1 lezen, gemiddeld genomen vijf respectievelijk acht fouten maken op deze kaarten.

Bij de kaarten van niveau 3 valt het hoge aantal fouten op van de leerlingen die lezen op AVI-niveau 1: 13 respectievelijk 12 fouten (zie tabellen 2.7 en 2.9). Voor kaart 4A en 4B vonden we een lineaire trend; voor de overige niveaus vonden we geen significante verschillen tussen leerlingen met een verschillend AVI-niveau, met uitzondering van de kaarten 8A en 8B.

Vrij snel na de try-out bleek dat we voor de normering en validering van de toetskaarten konden aanhaken bij het longitudinale onderzoek naar schoolvorderingen van de vakgroep Onderwijskunde Nijmegen (Aarnoutse, Van Leeuwe, Voeten, Van Kan & Oud, 1996). In de planning van dat project dienden de toetskaarten in maart 1992 voor de eerste maal afgenomen te worden. Dit was tevens de enige mogelijkheid in dat project om leerlingen uit groep 3 te toetsen. Doordat we zo snel moesten beginnen met de toetsafnames was het niet meer mogelijk om nog nadere analyses uit te voeren op de gegevens van de tweede try-out en op basis daarvan de toetskaarten eventueel nog te veranderen. Daardoor zijn de toetskaarten nauwelijks gewijzigd naar aanleiding van de tweede try-out. Alleen in kaart 1A is het woord 'por' vervangen door het woord 'duw', omdat bleek dat de kinderen het eerste woord niet kenden.

In bijlage 2 hebben we ter illustratie twee toetskaarten afgebeeld. In tabel 2.10 geven we een overzicht van de kenmerken van de uiteindelijke toetskaarten: het aantal zinnen, het aantal woorden, de gemiddelde woordlengte (WL), de gemiddelde zinslengte (ZL) en de Leesindex A. De waarde van de Leesindex neemt duidelijk af naarmate het AVI-niveau toeneemt. Vergelijking met de criteria voor de AVI-niveaus (bijlage 1) laat zien dat de waarden van de Leesindex voor de niveaus 5 tot en met 9 telkens binnen de voorgeschreven range vallen: 95 voor niveau 5 (range 99-94), 91-92 voor niveau 6 (range 93-89), 87-85 voor niveau 7 (range 88-84), 80-81 voor niveau 8 (range 83-79) en 76-75 voor niveau 9 (range 78-74). Het verschil in Leesindex tussen de opeenvolgende niveaus (vanaf niveau 5) bedraagt circa 4 punten.

Tabel 2.10: Overzicht gegevens nieuwe AVI-toetskaarten

	zinnen	woorden	letter- grepen	ZL [*]	WL ^{**}	Lees- index A
KAARTEN NIVEAU 1						
A naar het bos	21	104	104	5,0	100	118
B mijn zus leent mij een boek	19	104	104	5,5	100	117
KAARTEN NIVEAU 2						
A gaat de bal door de ruit?	24	129	135	5,4	105	114
B jou knijp ik nog wel!	24	124	131	5,2	106	114
KAARTEN NIVEAU 3						
A De hond van Ans is niet bang	20	132	153	6,6	116	105
B De goudvis	21	143	166	6,8	116	104
KAARTEN NIVEAU 4						
A Op zoek naar..	21	149	181	7,1	121	100
B Een ongeluk	25	156	187	6,2	120	103
KAARTEN NIVEAU 5						
A De brommer van mijn broer	22	164	210	7,5	128	95
B Joost spaart schelpen	21	154	197	7,3	128	95
KAARTEN NIVEAU 6						
A De spreekbeurt	20	150	201	7,5	134	91
B Op het toneel	20	151	199	7,6	132	92
KAARTEN NIVEAU 7						
A Verdwaald	24	206	282	8,6	137	87
B Naar de top	22	199	275	9,0	138	85
KAARTEN NIVEAU 8						
A Koeien op de heide	21	204	293	9,7	144	80
B Een moderne grap	19	182	260	9,6	143	81
KAARTEN NIVEAU 9						
A De ruimtevlucht	22	230	340	10,5	148	76
B De lucht in	20	226	330	11,3	146	75

NB ZL^{*} (zinslengte) = aantal woorden/aantal zinnenWL^{**} (woordlengte) = 100*aantal lettergrepen/aantal woorden

2.3 De onderzoeksvragen

In dit hoofdstuk hebben we kort weergegeven op welke wijze de nieuwe AVI-toetskaarten tot stand zijn gekomen. Tevens zijn we nader ingegaan op de indeling in de negen AVI-niveaus, waarop de toetskaarten gebaseerd zijn. In de volgende twee hoofdstukken zullen de schaling en de normering respectievelijk de validering van de nieuwe toetskaarten aan de orde komen. Daarbij staan de volgende onderzoeksvragen centraal:

- 1 Lopen de nieuwe AVI-toetskaarten op in moeilijkheidsgraad?
- 2 Wat is de betrouwbaarheid van de AVI-toetskaarten?
- 3 Hoe is de verdeling van de AVI-niveaus over de basisschoolperiode?
- 4 Kunnen de A- en de B-versie van de AVI-toets als paralleltests worden beschouwd?
- 5 Wat is de relatie tussen de AVI-toets en de Eén-Minuut-Test (EMT)?
- 6 Wat is de relatie tussen de AVI-toets en toetsen voor begrijpend lezen en woordenschat?

De onderzoeksvragen 1 tot en met 3 komen aan de orde in hoofdstuk 3. De eerste stap in het daar gepresenteerde onderzoek is het construeren van één schaal voor het leesniveau van de leerling. Elke toetskaart vormt een onderdeel van de schaal, maar beslaat niet de gehele schaal. De bedoeling is uiteindelijk dat het niet nodig zal zijn om alle kaarten bij een leerling af te nemen, maar slechts enkele. De prestaties van de leerling op dat beperkte aantal kaarten moet dan voldoende zijn om het leesniveau vast te kunnen stellen. Daarvoor is het belangrijk dat de kaarten geleidelijk oplopen in moeilijkheidsgraad (onderzoeksvraag 1). Voor het hanteren van de toetskaarten in de praktijk zal de schaal voor de leestijden in 11 stukken verdeeld worden, van 'lager dan AVI-niveau 1' tot en met 'hoger dan AVI-niveau 9'. De grenswaarden op deze schaal worden omgezet naar leestijden op de negen kaarten van elke versie. Tevens geldt per kaart een maximaal aantal toegestane fouten. Vervolgens zal nagegaan worden of de toets consistent meet (onderzoeksvraag 2). Tot slot wordt in hoofdstuk 3 nagegaan hoe de verdeling is van de AVI-niveaus op de diverse tijdstippen in het basisonderwijs (onderzoeksvraag 3). Dit vormt tevens de normering van de AVI-toetskaarten.

In hoofdstuk 4 worden de onderzoeksvragen 4 tot en met 6 beantwoord. Onderzoeksvraag 4 betreft feitelijk nog de betrouwbaarheid van de toets, maar wordt in dit hoofdstuk beantwoord omdat deze vraag met dezelfde onderzoekstechniek wordt beantwoord als de vragen 5 en 6. Deze twee vragen betreffen de validiteit van de AVI-toetskaarten. De AVI-toets is bedoeld voor het bepalen van de technische leesvaardigheid van de leerlingen. In het validiteitsonderzoek zullen we nagaan of de resultaten op de AVI-toets voldoen aan wat we op basis van de theorie mogen verwachten.

Op de eerste plaats mogen we verwachten dat de samenhang van de AVI-toets met een andere toets voor technisch lezen, in dit geval de Eén-Minuut-Test, hoog is (onderzoeksvraag 5). Het belangrijkste verschil tussen beide toetsen is het toetsmateriaal: in de AVI-toets wordt gebruik gemaakt van teksten, in de EMT van losse woorden.

De samenhang met toetsen voor begrijpend lezen en woordenschat (onderzoeksvraag 6), die andere aspecten van het lezen meten, dient minder hoog te zijn. Op basis van de theorie gepresenteerd in paragraaf 1.4.1 verwachten we dat een toename van de technische leesvaardigheid een positief effect heeft op de vaardigheid in begrijpend lezen. Daarnaast verwachten we dat de samenhang tussen de vaardigheid in technisch lezen en begrijpend lezen het grootst is aan het begin van het leesonderwijs en afneemt naarmate de vaardigheid van de leerlingen toeneemt.

3 SCHALING EN NORMERING

In dit hoofdstuk staan de schaling en normering van de nieuwe AVI-toetskaarten centraal. We proberen de kaarten te schalen op basis van empirische gegevens, waarbij we uitgaan van een 'latente trek'-model van de leestijden. We zullen vervolgens nagaan in hoeverre de moeilijkheidsgraad van de toetskaarten, zoals bepaald met behulp van dit model, overeenkomt met de indeling naar AVI-niveau. Vervolgens bepalen we de grenswaarden, dat wil zeggen de indelingscriteria voor de AVI-niveaus (vergelijk tabel 1.3 in paragraaf 1.1.2). Evenals bij de oude AVI-kaarten dienen deze grenswaarden om aan te kunnen geven of een leerling in staat is om leesstof van een bepaalde moeilijkheidsgraad te lezen.

De volgende onderzoeksvragen komen in dit hoofdstuk aan de orde:

- 1 Lopen de nieuwe AVI-toetskaarten op in moeilijkheidsgraad?
- 2 Wat is de betrouwbaarheid van de AVI-toetskaarten?
- 3 Hoe is de verdeling van de AVI-niveaus over de basisschoolperiode?

Het hoofdstuk is als volgt opgebouwd. De onderzoeksopzet staat beschreven in paragraaf 3.1. In paragraaf 3.2 bespreken we de schaling van de toetskaarten (onderzoeksvraag 1) en de wijze waarop de grenswaarden voor de kaarten tot stand zijn gekomen. Vervolgens wordt in paragraaf 3.3 de betrouwbaarheid van de uitkomsten besproken (onderzoeksvraag 2). Paragraaf 3.4 bevat de normering van de AVI-toets (onderzoeksvraag 3). Tot slot worden in paragraaf 3.5 enkele conclusies getrokken ten aanzien van de drie onderzoeksvragen.

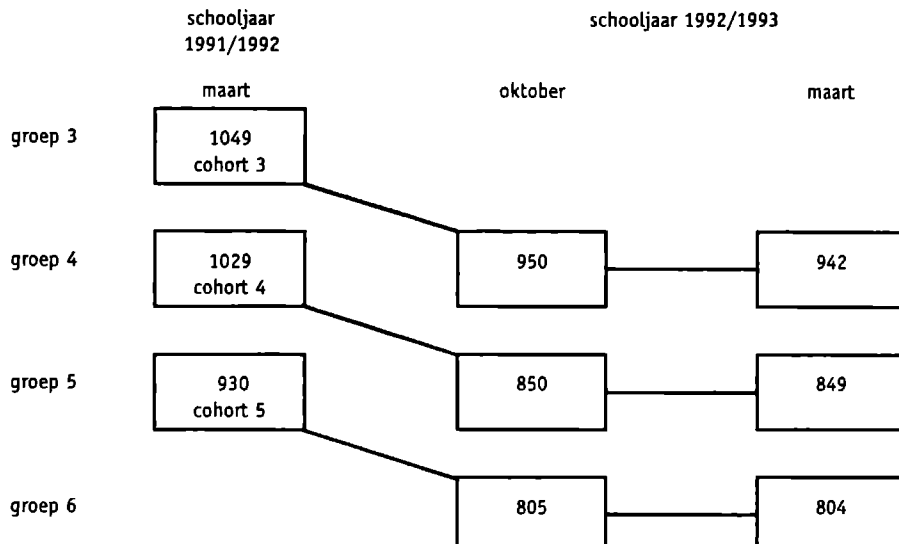
3.1 De onderzoeksopzet

3.1.1 De steekproef

Voor het verkrijgen van een betrouwbare en valide schaling en normering zijn de kaarten afgenomen bij een steekproef uit de doelgroep: leerlingen van groep 3 tot en met 6 van de basisschool. Voor het normeren van de AVI-toets is aangesloten bij het longitudinale onderzoek naar schoolvorderingen in het basisonderwijs van de Vakgroep Onderwijskunde Nijmegen (Aarnoutse, Van Leeuwe, Oud & Voeten, 1991). In dat onderzoek, gestart in september 1991 met een landelijke steekproef van 39 scholen, worden drie cohorten leerlingen gevolgd gedurende de basisschoolperiode: leerlingen die in september 1991 aanwezig waren in respectievelijk groep 3, 4 en 5 van het basisonderwijs. We duiden deze cohorten in het vervolg van dit proefschrift aan als cohort 3, 4 en 5.

Overigens is deze steekproef als gevolg van uitval niet volledig representatief voor alle Nederlandse basisscholen: scholen in grote steden met veel allochtone leerlingen en Nederlandse achterstandsleerlingen zijn ondervertegenwoordigd (Aarnoutse, Van Leeuwe, Voeten, Van Kan & Oud, 1996).

Voor de schaling en normering van de AVI-toets hebben we gebruik gemaakt van de gegevens verzameld op drie tijdstippen: maart 1992, oktober 1992 en maart 1993. In figuur 3.1 staat weergegeven bij hoeveel leerlingen per groep per tijdstip de AVI-toets is



Figuur 3.1: Samenstelling van de cohorten van het normeringsonderzoek (aantallen leerlingen)

afgenomen. Door de opzet van het onderzoek beschikken we voor groep 4, 5 en 6 over gegevens op twee tijdstippen in het schooljaar (oktober en maart) en voor groep 3 over gegevens op één tijdstip (maart). Het grote verschil in het aantal leerlingen tussen maart en oktober 1992 wordt veroorzaakt doordat enkele scholen zijn uitgevallen. In maart 1992 is de AVI-toets op 35 scholen afgenomen, in oktober 1992 en maart 1993 op 32 scholen, waarvan op één school de toets niet in maart 1992 was afgenomen.

3.1.2 De dataverzameling

In maart 1992 werd bij alle leerlingen van groep 3, 4 en 5 zowel de A- als de B-versie van de nieuwe AVI-toets afgenomen. De twee versies werden afgenomen met een tussenperiode van minimaal een dag en, indien mogelijk, maximaal een week. Op de ene helft van de scholen werd gestart met versie A, op de andere helft met versie B. In oktober 1992 is bij alle leerlingen van groep 4, 5 en 6, om de belasting van de scholen en de leerlingen te beperken, alleen de A-versie afgenomen. In maart 1993 werden bij alle leerlingen van groep 4, 5 en 6 weer beide versies van de AVI-toets afgenomen, om gegevens te verkrijgen over de vergelijkbaarheid van de A- en B-versie.

De afname van de AVI-toets in maart 1992 geschiedde in principe door de school zelf. Op enkele grote scholen is ondersteuning geboden door medewerkers van het KPC. De afnames in oktober 1992 en maart 1993 zijn grotendeels uitgevoerd door PABO-studenten. Zowel de leerkrachten als de PABO-studenten zijn uitgebreid geïnstrueerd over de wijze waarop de toets afgenomen diende te worden (KPC, 1992a). Voor de instructie is

een speciale videoband ontwikkeld waarin de afnameprocedure voor het onderzoek wordt uitgelegd en geoefend (KPC, 1992b).

De verkorte versie van de afname-instructie is opgenomen in bijlage 7. De wijze van afname van elke toetskaart is zoveel mogelijk gelijk gehouden aan de afnameprocedure zoals die voor de oude en de nieuwe AVI-toetskaarten geldt (Visser, Van Laarhoven & Ter Beek, 1996) en zoals die in de try-out is gehanteerd. Dezelfde criteria zijn gehanteerd voor het bepalen wat goed en fout is (zie ook paragraaf 1.1.2). Tijdens de afname is wederom gebruik gemaakt van optisch leesbare scoreformulieren (zie bijlage 5), waarop de fout gelezen woorden aangestreept konden worden. Evenals bij de tweede try-out (zie paragraaf 2.2.4) werd tevens het type fout genoteerd: fout gelezen, spellend gelezen, overgeslagen of omgewisseld, een woord tussengevoegd en voorgezegd. Nadat de leerling de gehele kaart had gelezen, werd op het formulier tevens de leestijd genoteerd. De leerlingen kregen ruim de tijd voor het lezen van de toetskaarten; in principe werd gestopt wanneer het lezen van een kaart langer dan vijf minuten duurde.

Het verschil met de afnameprocedure in de handleiding van de toets (Visser et al, 1996) is dat niet op basis van de prestaties van de leerling (aantal fouten en leestijd) bepaald kon worden welke toetskaarten de leerling achtereenvolgens diende te lezen. De criteria daarvoor wilden we immers met dit onderzoek vaststellen. Van te voren moesten we daarom bepalen welke kaarten de leerlingen voorgelegd kregen.

De meest ideale situatie voor de schaling van de kaarten zou zijn die waarin leerlingen, ongeacht hun leesniveau, alle AVI-kaarten lezen. Op deze wijze kan een zo goed mogelijk inzicht verkregen worden in de moeilijkheidsgraad van de kaarten. Deze opzet was echter niet haalbaar. Voor de zwakkere lezers is het moeilijk, zo niet onmogelijk om tot en met kaart 9 te lezen. En voor alle leerlingen geldt dat de tijdsbelasting te groot zou zijn. Immers, uitgaande van gemiddeld 3,5 minuut per kaart zou de afname van de A- en B-versie ruim 1 uur per leerling vergen.

Daarom is gekozen voor een opzet waarbij leerlingen in principe drie opeenvolgende kaarten van een versie lazen die qua moeilijkheid in de buurt van hun leesniveau lagen. Dat leesniveau is voor maart 1992 ingeschat met behulp van de score op de Eén-Minuut-Test (EMT) (Brus & Voeten, 1973) voor de leerlingen uit groep 4 en 5 (zie bijlage 8). Voor de leerlingen van groep 3 is het leesniveau ingeschat door de groepsleerkracht, aangezien de EMT voor deze leerlingen op dat tijdstip niet geschikt is. Daarbij konden ze eventueel gebruikmaken van de score op de Caesar-EMT (Mommers, 1983; zie bijlage 8). In oktober 1992 en maart 1993 is het leesniveau ingeschat op basis van de prestaties op de nieuwe AVI-toets van de voorgaande periode.

Iedere leerling kreeg in principe drie toetskaarten van een versie voorgelegd: één kaart op het ingeschatte niveau, één kaart onder en één kaart boven dat niveau. Een uitzondering hierop vormden leerlingen met ingeschat AVI-niveau 1 en 9. Zij lazen twee kaarten van een versie: de kaart van het ingeschatte niveau en de kaart van een niveau hoger, respectievelijk lager.

Om te voorkomen dat een leerling alleen maar te moeilijke of te makkelijke kaarten zou lezen (immers de indeling is gebaseerd op een ingeschat leesniveau) zijn de volgende regels gehanteerd. Als een leerling meer dan drie minuten nodig had om een kaart te lezen, werden geen kaarten meer aangeboden van een hoger niveau. Wel werd een kaart

van een lager niveau dan de eerst aangeboden gelezen, totdat in totaal drie kaarten van de betreffende versie waren gelezen. Leerlingen die hun laatste (=hoogste) kaart foutloos binnen 1'20" lazen, kregen nog een extra kaart van één niveau hoger aangeboden, tenzij de kaart van niveau 9 al was gelezen. In tabel 3.1 staat per tijdstip vermeld hoeveel leerlingen de verschillende combinaties van niveaus van kaarten hebben gelezen.

Tabel 3.1: Gelezen combinaties van kaarten per versie per tijdstip (aantal leerlingen)

combinatie van kaarten	maart 1992		oktober 1992	maart 1993	
	versie A	versie B	versie A	versie A	versie B
1	123	96	14		
1-2	762	764	235	63	66
1-2-3	268	277	282	130	130
2-3-4	192	185	364	301	299
3-4-5	209	207	159	290	299
4-5-6	231	233	200	210	233
5-6-7	206	211	218	236	249
6-7-8	311	312	242	248	248
7-8-9	321	322	282	266	276
8-9	341	337	494	645	668
oveng	26	13	115	92	85

Uit de tabel kan afgelezen worden dat in maart 1992 een deel van de leerlingen alleen een toetskaart van niveau 1 heeft gelezen. Deze leerlingen, allen uit groep 3, bleken al veel moeite te hebben met het lezen van de kaart van niveau 1, zodat tijdens de afname besloten is hen niet ook nog een kaart van niveau 2 aan te bieden. Op de andere tijdstippen komt dit niet of nauwelijks meer voor.

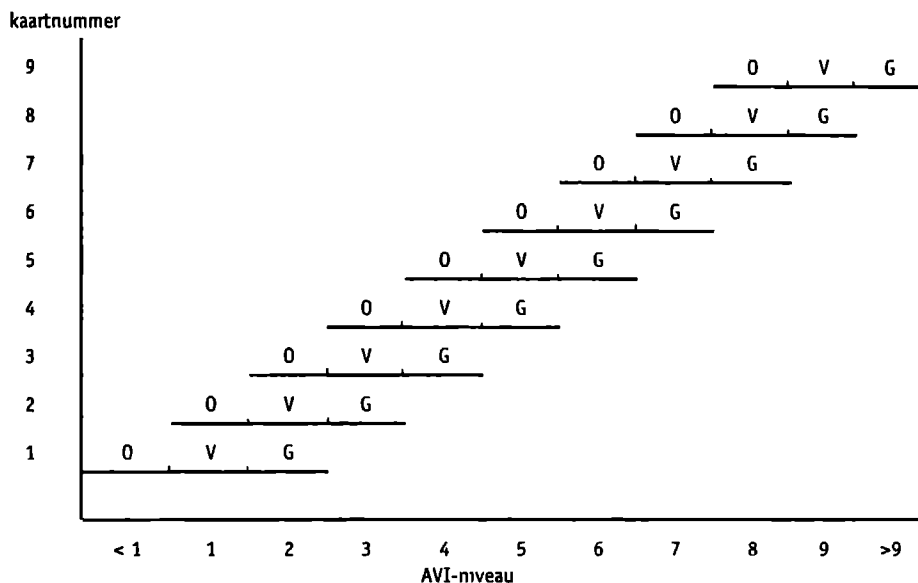
In bijlage 9 staan de beschrijvende statistieken van de leestijden en fouten per toetskaart. Voor maart 1992, oktober 1992 en maart 1993 zijn per groep een tabel voor de leestijden en een tabel voor het aantal fouten opgenomen.

3.2 Het bepalen van de grenswaarden

3.2.1 Uitgangspunten

In paragraaf 1.1.2 hebben we al aangegeven dat de leestijd en het aantal fouten het AVI-niveau van de leerling bepalen. Het doel van het schalings- en normeringsonderzoek was het bepalen van de grenswaarden: bij welk aantal fouten en welke leestijd behoort

welk AVI-niveau (vergelijk tabel 1.3)? Elk niveau, en dus ook de bijbehorende toetskaart, kan gezien worden als een trapje op een ladder. Een toetskaart discrimineert niet over het gehele continuüm van technische leesvaardigheid, maar heeft betrekking op een stukje ervan. In figuur 3.2 wordt een en ander gevisualiseerd.



Figuur 3.2: Correspondentie tussen AVI-niveaus en (de prestaties op) de toetskaarten (O=onvoldoende, V=voldoende, G=goed)

Per kaart worden drie prestatieniveaus onderscheiden: onvoldoende (O), voldoende (V) en goed (G). Op basis van deze prestatieniveaus worden de AVI-niveaus toegekend. Voldoende betekent dat de leerling leest op het niveau van de betreffende toetskaart. De implicatie hiervan is dat de leerling teksten aan kan die dezelfde moeilijkheidsgraad hebben als de toetskaart. Onvoldoende geeft aan dat dat niveau nog niet wordt beheerst, goed betekent dat het kind het niveau volledig beheerst. In dat laatste geval kan het kind ook teksten aan van een hoger niveau. Welk niveau dat precies is, wordt bepaald door het kind toetskaarten te laten lezen van een hoger niveau.

Per kaart dienen we voor de leestijd en het aantal fouten twee grenswaarden te bepalen, namelijk voor de overgang van onvoldoende naar voldoende en voor de overgang van voldoende naar goed. De grenswaarden die uiteindelijk het leesniveau bepalen, dienen zo gekozen te worden dat de uitkomsten van verschillende kaarten zo consistent mogelijk zijn. Naar verwachting zal nooit volledige consistentie bereikt kunnen worden. Zo zullen toevalligheden de prestaties van de leerlingen beïnvloeden waardoor een leerling de ene keer wat beter presteert dan de andere keer. Als gevolg daarvan zal ook het toegekende AVI-niveau van de leerling kunnen verschillen. We zullen daarom de criteria probabilistisch formuleren:

- 1 de kans op een verschillend resultaat voor twee kaarten van hetzelfde niveau moet zo klein mogelijk zijn;
- 2 de kans op een voldoende of goed resultaat op de hoogste van twee kaarten, gegeven een onvoldoende of voldoende resultaat op de lagere kaart, moet zo klein mogelijk zijn;
- 3 de kans op een onvoldoende resultaat op de hoogste van twee kaarten, gegeven een goed resultaat op de lagere kaart, moet zo klein mogelijk zijn.

Kijkend naar figuur 3.2 betekent het eerste criterium bijvoorbeeld dat wanneer een leerling op kaart 3 versie A de score voldoende behaalt, de kans op een andere score dan voldoende op kaart 3 versie B zo klein mogelijk moet zijn.

Het tweede criterium houdt in dat bijvoorbeeld bij een voldoende prestatie op een kaart van niveau 3 de kans op een voldoende of goede prestatie op een kaart van niveau 4 klein is. Immers uit de figuur is eenvoudig af te leiden dat een voldoende prestatie op kaart 3 correspondeert met een onvoldoende prestatie op kaart 4.

Het derde criterium, ten slotte, betekent dat bijvoorbeeld bij een goede prestatie op kaart 2 de kans op een onvoldoende op kaart 3 klein is. Goed op kaart 2 correspondeert in de figuur immers met voldoende of eventueel goed op kaart 3.

Bij het lezen van de AVI-kaarten zijn zowel de tijd die een leerling nodig heeft om een kaart te lezen als het aantal fout gelezen woorden van belang als indicatoren van de technische leesvaardigheid van de leerling. Momenteel is geen geschikt model beschikbaar dat deze beide indicatoren in hun onderlinge relatie beschrijft. Rasch (1960) presenteerde een Poisson-model voor het aantal fout gelezen woorden bij het hardop lezen van een tekst. Tevens presenteerde hij een model voor de leestijden. Dit laatste model hebben we voor de schaling van de AVI-toetskaarten als uitgangspunt genomen, aangezien de leessnelheid de belangrijkste indicator is voor de mate waarin de technische leesvaardigheid geautomatiseerd is (zie paragraaf 1.2.4). De nauwkeurigheid waarmee woorden worden gelezen is al snel groot waardoor het aantal fouten niet meer voldoende differentieert tussen leerlingen.

Onze aanpak bestaat er daarom uit om een schaal voor de technische leesvaardigheid te bepalen op basis van de leestijden. De AVI-niveaus worden vervolgens gedefinieerd door het opdelen van deze schaal in 11 stukken, van 'lager dan AVI-niveau 1' tot en met 'hoger dan AVI-niveau 9'. Wanneer een kind echter 'te veel' fouten heeft gemaakt dan dient zijn leesniveau in benedenwaartse richting te worden gecorrigeerd. De leessnelheid mag niet bereikt worden ten koste van de nauwkeurigheid van het lezen. In deze redenering hebben we dus eerst een model nodig dat ons in staat stelt om de leestijden benodigd voor de AVI-kaarten te schalen (paragraaf 3.2.2). Vervolgens dienen op deze schaal de grenzen voor de AVI-niveaus te worden gemarkeerd (paragraaf 3.2.3) en ten slotte hebben we een criterium nodig voor het bepalen van het maximaal aantal toegestane fouten (paragraaf 3.2.4). Op basis hiervan bepalen we de uiteindelijke grenswaarden (paragraaf 3.2.5).

3.2.2 Schaling op basis van de leestijden

Rasch (1960) formuleerde zijn model voor leestijden vanuit de assumptie dat gemakkelijke en moeilijke tekstdelen (woorden) gelijkmatig gespreid zijn over de hele tekst en

dat de leerling de hele tijd blijft lezen met dezelfde gemiddelde snelheid. Bovendien nam hij aan dat op elk willekeurig moment tijdens het lezen de leessnelheid onafhankelijk is van het aantal tot op dat moment gelezen woorden en ook van de wijze waarop die woorden gelezen werden. Op basis van deze assumpties kon Rasch de kans specificeren om een tekst bestaande uit m woorden binnen een bepaald tijdsbestek hardop te lezen. Deze kans volgt een gamma-verdeling die bepaald wordt door het aantal woorden m , waaruit de tekst bestaat en door de te schatten intensiteitsparameters λ_y :

$$p(t_y | m_y, \lambda_y) = \frac{1}{(m_y - 1)!} \lambda_y^{m_y} t_y^{m_y - 1} \exp(-\lambda_y t_y) \quad (3.1)$$

Hierbij is t_y de tijd (in seconden) die leerling i nodig heeft om toetskaart j hardop te lezen. De tekstlengte m_y is gegeven: het aantal woorden van tekst j . Hierbij wordt dus de assumptie gemaakt dat de woorden zinvolle eenheden zijn waarin een tekst kan worden opgesplitst, hetgeen zeer plausibel lijkt.

Rasch stelde voor om de intensiteitsparameters te beschouwen als het produkt van de moeilijkheid van de tekst en de leesvaardigheid van de leerling:

$$\lambda_y = \theta_i \epsilon_j \quad (3.2)$$

Hierbij is θ_i de leesvaardigheid van een willekeurige leerling i , en ϵ_j de moeilijkheid van toetskaart j . In de intensiteitsparameter λ_y worden de moeilijkheidsgraad van de toetskaart j en de leesvaardigheid van leerling i aan elkaar gerelateerd. Deze twee samen bepalen hoeveel tijd leerling i nodig heeft om de toetskaart j te lezen.

Jansen (1997) heeft het model van Rasch verder uitgewerkt. Zij veronderstelt een 'mixed' model met ϵ_j als 'fixed' en de bekwaamheidsparameters θ_i als 'random'. Zij neemt aan dat de bekwaamheidsparameters, die altijd een positieve waarde moeten hebben, een gamma-verdeling volgen met twee parameters die het gemiddelde en de variantie van de verdeling vastleggen. Zij ontwikkelde 'marginal maximum likelihood'-schatters voor de moeilijkheidsgraden en voor de twee parameters van de leesvaardigheidsverdeling. Daarbij werd rekening gehouden met de situatie dat een leerling steeds slechts een deel van de AVI-kaarten leest (zie paragraaf 3.1.2), waarbij het dan bovendien ook nog zo is dat de toedeling van AVI-kaarten aan leerlingen volgens een systematische procedure gebeurt. Immers, welke kaarten men een leerling laat lezen wordt bepaald door externe informatie (prestaties op de EMT of de inschatting van de leerkracht) over de leesvaardigheid van die leerling (zie paragraaf 3.1.2). Voor verdere details ten aanzien van model en schattingsprocedure wordt verwezen naar Jansen (1997).

Het Rasch-model is door Jansen toegepast op de AVI-toetsgegevens van maart 1992. Toepassing van het model leverde gegevens op over de moeilijkheidsgraad van de 18 toetskaarten. Bij de ontwikkeling van de AVI-kaarten, zoals eerder vermeld, zijn de teksten reeds qua moeilijkheid geordend op basis van de AVI-niveaucriteria (zie paragraaf

2.1.2 en bijlage 1). In het model wordt er echter vanuit gegaan dat zowel de moeilijkheidsgraad van de teksten als de leesvaardigheid van de leerlingen uit de data geschat moeten worden. Op deze wijze kunnen we de constructie volgens de AVI-niveaucriteria empirisch controleren op basis van de leestijden (onderzoeksvraag 1). In hoofdstuk 2 hebben we onze twijfels uitgesproken over de betrouwbaarheid van de indeling van de vijf hoogste AVI-niveaus op grond van Leesindex A. Op grond van de schaling kunnen we nagaan of de volgorde volgens de AVI-niveau-indeling overeenkomt met de volgorde in moeilijkheidsgraad geschat op basis van het door Jansen aangepaste Rasch-model voor leestijden. Tabel 3.2 geeft de geschatte moeilijkheidsgraden van de toetskaarten weer. Hoe lager $\hat{\epsilon}$, hoe moeilijker de toetskaart.

Tabel 3.2: Leesindex A, geschatte moeilijkheidsparameters ($\hat{\epsilon}$) van de toetskaarten

kaart	versie A		versie B	
	leesindex-A	$\hat{\epsilon}$	Leesindex-A	$\hat{\epsilon}$
1	118	1,4546	117	1,2937
2	114	1,2637	114	1,2401
3	105	0,9733	104	1,0023
4	100	0,9440	103	1,0533
5	95	0,9618	95	0,9180
6	91	0,9626	92	0,9529
7	87	0,8445	85	0,8579
8	80	0,8231	81	0,8638
9	76	0,8037	75	0,7866

Het verloop van de moeilijkheidsparameters blijkt niet altijd conform de veronderstelde moeilijkheidsgraden van de teksten gebaseerd op de indeling in AVI-niveaus. In versie A is kaart 4 iets moeilijker dan kaart 5 en 6, welke laatste twee ongeveer even moeilijk zijn. In versie B is kaart 4 makkelijker dan kaart 3, kaart 6 makkelijker dan kaart 5 en is kaart 8 ongeveer even moeilijk als kaart 7. Ook verschilt de moeilijkheid van kaarten die bedoeld zijn voor hetzelfde niveau (bijvoorbeeld kaart 1A en 1B). Naast het feit dat de volgorde in moeilijkheidsgraad niet overal klopt, zien we tevens dat afstanden tussen de moeilijkheidsgraden van de kaarten niet gelijk zijn. De afstand tussen kaart 1A, 2A en 3A is veel groter dan die tussen de kaarten van de hogere niveaus. Voor versie B geldt dat het verschil in moeilijkheid tussen 1B en 2B vrij klein is, terwijl het verschil tussen 2B en 3B van dezelfde orde is als het verschil tussen de eerste drie kaarten van de A-versie. Tot slot geldt dat de moeilijkheidsgraden van nogal wat kaarten dichtbij elkaar liggen. Zo liggen die van kaart 3A, 4A, 5A, 5B, 6A en 6B alle tussen 0,92 en 0,97 en die van 7A, 7B, 8A, 8B, 9A en 9B tussen 0,79 en 0,86.

Deze empirische controle van de moeilijkheidsgraad van de toetskaarten wijst erop dat de indeling volgens de AVI-niveaucriteria niet strikt hiërarchisch is. Met name in de hogere niveaus, waar de Leesindex A het enige criterium is, wijkt de a priori volgorde af van de empirische. Wanneer we bijvoorbeeld kaart 4A en kaart 6A vergelijken, dan zien we dat de leesindex voor deze kaarten 100 respectievelijk 91 bedraagt. De geschatte moeilijkheidsgraad bedraagt voor deze twee kaarten 0,9440 respectievelijk 0,9626, waarbij kaart 4A ongeveer even moeilijk is of zelfs iets moeilijker lijkt te zijn dan 6A. Naast woord- en zinslengte (opgenomen in de Leesindex A) spelen andere factoren een rol die de moeilijkheid van de tekst bepalen. Wanneer we de inhoud van deze twee kaarten bekijken, blijkt dat de tekst van 6A (over het houden van een spreekbeurt) dichter bij de belevingswereld van kinderen ligt dan die van 4A (over gaswinning). Hieruit kunnen we concluderen dat de Leesindex A geen voldoende criterium is voor de daadwerkelijke moeilijkheid van een tekst, hetgeen ook door anderen reeds is aangetoond. Daarbij moeten we ook opmerken dat de verschillen in leesindex tussen AVI-niveaus klein zijn (zie ook paragraaf 2.1.2). De hogere AVI-niveaus beslaan ieder een zeer beperkt aantal waarden op de leesindex-schaal. Bijvoorbeeld: niveau 6 betreft de waarden 93-99 als leesindex, niveau 7 betreft de waarden 88-94. Deze kleine verschillen op de Leesindex A zijn blijkbaar niet voldoende om waarneembare verschillen in moeilijkheid te onderscheiden.

Wat betekent nu het feit dat de veronderstelde volgorde in moeilijkheid niet overeenkomt met de empirisch vastgestelde moeilijkheidsgraad? Eigenlijk zou het gewenst zijn enkele teksten te herschrijven danwel nieuwe teksten te schrijven en deze vervolgens opnieuw aan een schalingsonderzoek te onderwerpen. Dit was echter niet mogelijk, gezien de daarvoor benodigde tijd en kosten. Het is ook niet mogelijk om de kaarten simpelweg om te draaien, zodat de teksten wel duidelijk toenemen in moeilijkheidsgraad. Immers, de teksten zijn geordend volgens de AVI-niveau-criteria, volgens welke ook leesboeken ingedeeld worden op AVI-niveau. Voorlopig zullen we de kaarten handhaven zoals ze zijn. De gevonden resultaten geven wel duidelijk aan dat in de praktijk bij het kiezen van leesstof het AVI-niveau (en de Leesindex A) niet als enig en absoluut criterium gehanteerd mag worden.

Met behulp van de moeilijkheidsparameters $\hat{\theta}$ van de toetskaarten en de geobserveerde leestijden kan voor iedere leerling de vaardigheidsparameter θ worden geschat. Daarvoor is de volgende formule gebruikt:

$$\hat{\theta}_{gi} = \frac{\sum_j v_{gj} m_j}{\sum_j \hat{\theta}_j v_{gj} t_{gij}} \quad (3.3)$$

Hierbij geeft de indicatorvariabele v_{gj} aan of een leerling een toetskaart heeft gemaakt. Als de leerlingen in groep g (zie tabel 3.1) toetskaart j hebben gemaakt, geldt $v_{gj}=1$ en anders $v_{gj}=0$.

Deze formule kan ook worden toegepast op de toetsgegevens van de leerlingen verkregen bij de metingen van oktober 1992 en maart 1993. Op basis van de leestijden op de toetskaarten die een leerling heeft gelezen, is zijn theta-waarde op de leesvaardigheids-schaal geschat. In tabel 3.3 staan per cohort per tijdstip het gemiddelde en de standaardafwijking van θ weergegeven.

Tabel 3.3: Gemiddelden en standaardafwijkingen op de theta-schaal voor leestijden per cohort per tijdstip

	cohort 3			cohort 4			cohort 5		
	Gemid- delde	Standaard- afwijking	N	Gemid- delde	Standaard- afwijking	N	Gemid- delde	Standaard- afwijking	N
maart groep 3	0,70	0,40	1040						
oktober groep 4	1,30	0,54	950						
maart groep 4	1,73	0,55	942	1,72	0,58	1029			
oktober groep 5				2,05	0,58	850			
maart groep 5				2,32	0,57	849	2,26	0,54	930
oktober groep 6							2,51	0,58	805
maart groep 6							2,69	0,55	804

Uit de tabel blijkt, conform de verwachtingen, een stijging van de gemiddelde technische leesvaardigheid in de tijd. De groei in technische leesvaardigheid is het sterkst in de beginperiode van het leesonderwijs (tussen maart groep 3 en oktober groep 4) en neemt af in de tijd. Dit komt overeen met bevindingen van ander onderzoek (Aarnoutse et al, 1996; Boland & Mommers, 1986; Seegers, 1985). De spreiding blijft vrij stabiel, met uitzondering van de meting in maart van groep 3. Voor twee van de zeven tijdstippen zijn gegevens van twee cohorten beschikbaar (maart groep 4 en maart groep 5). De gemiddelde schaalwaarden op die tijdstippen voor de verschillende cohorten zijn redelijk vergelijkbaar.

Voor de overige analyses die we in dit hoofdstuk bespreken, hebben we vervolgens één bestand gecreëerd voor de totale periode van maart groep 3 tot en met maart groep 6. Per leerling zijn toetsgegevens beschikbaar voor maximaal drie verschillende tijdstippen. We hebben het databestand zodanig gereorganiseerd dat elke meting van een leerling als een afzonderlijke 'case' wordt beschouwd. Voor twee tijdstippen (maart van groep 4 en groep 5) zijn gegevens beschikbaar van twee cohorten leerlingen. Om te voorkomen dat daardoor deze tijdstippen oververtegenwoordigd zijn in het totale bestand, hebben we aan deze 'cases' het gewicht 0,5 toegekend, terwijl de overige 'cases' het gewicht 1 meekregen. Tabel 3.4 geeft het aantal leerlingen per tijdstip weer, ongewogen en gewogen.

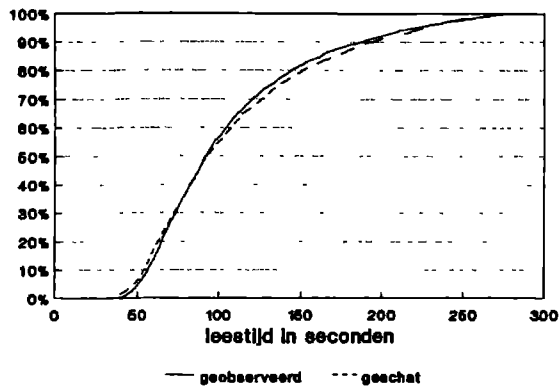
Tabel 3.4: Aantal leerlingen per tijdstip in het totale bestand

tijdstip	ongewogen N	gewogen N
maart groep 3	1040	1040
oktober groep 4	950	950
maart groep 4	1971	986
oktober groep 5	850	850
maart groep 5	1779	890
oktober groep 6	805	805
maart groep 6	804	804
totaal	8199	6325

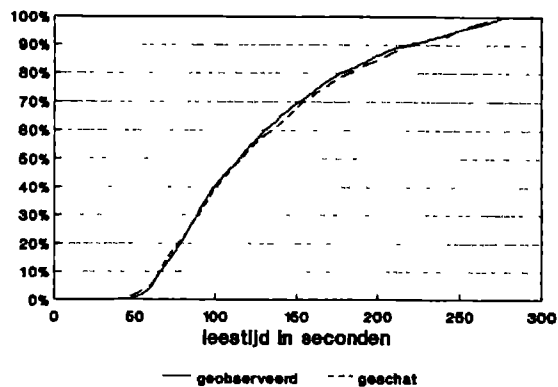
Alvorens over te gaan tot het bepalen van de grenswaarden op basis van de leestijden, hebben we enkele controles op het gebruikte model uitgevoerd. Analooq aan Jansen (1997) hebben we voor iedere toetskaart de cumulatieve verdeling bepaald van de geobserveerde en geschatte leestijden. Formule 3.4 geeft de verwachte leestijd op toetskaart j voor een leerling met geschatte vaardigheid $\hat{\theta}_g$.

$$E(t_{gy} | \hat{\theta}_g, \hat{e}_j) = \frac{m_j}{\hat{\theta}_g \hat{e}_j} \quad (3.4)$$

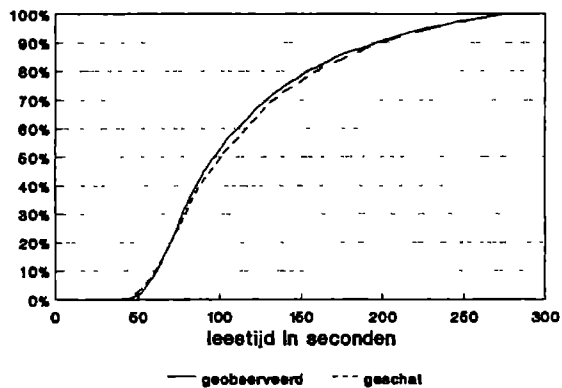
Voor iedere leerling i kan voor iedere toetskaart j de leestijd worden geschat door in deze formule de geschatte vaardigheid $\hat{\theta}_g$, de tekstlengte m_j en de geschatte moeilijkheidsparameter \hat{e}_j in te vullen.



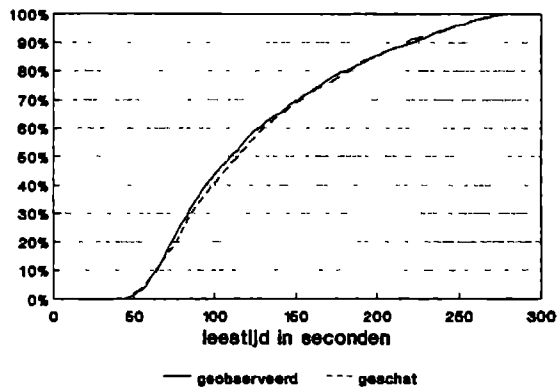
Kaart 2A



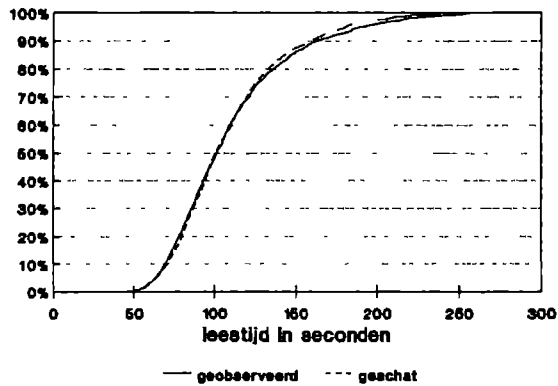
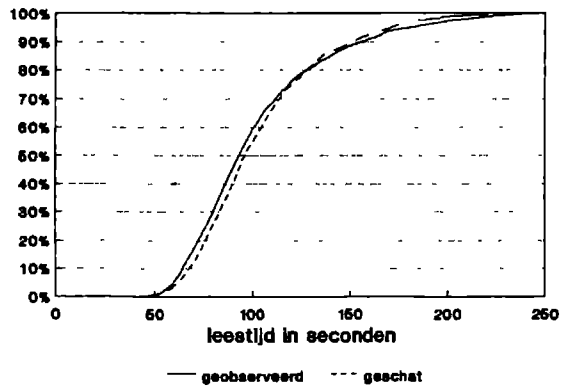
Kaart 2B



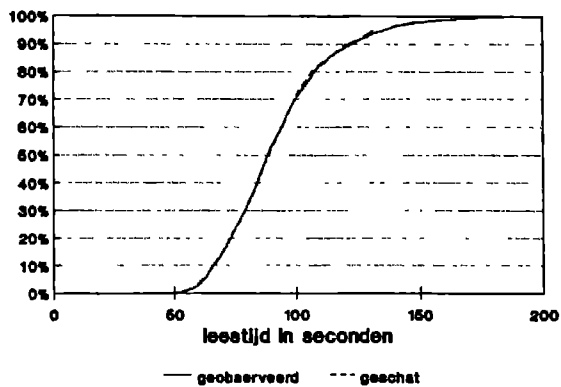
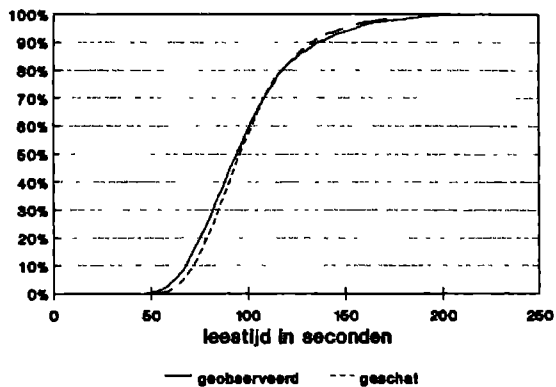
Kaart 3A



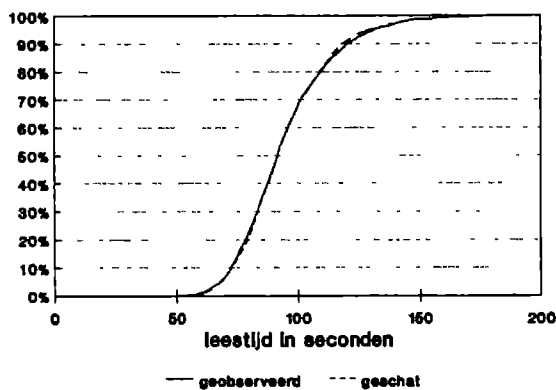
Kaart 3B



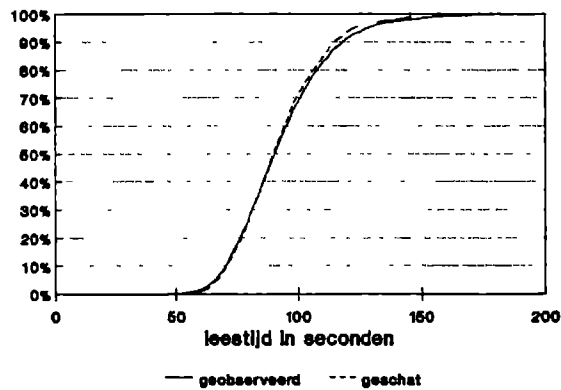
Figuur 3.3: Cumulatieve verdeling van geobserveerde en geschatte leestijden voor de



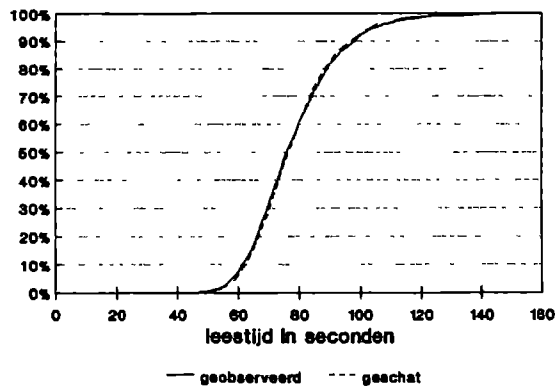
Kaart 5A



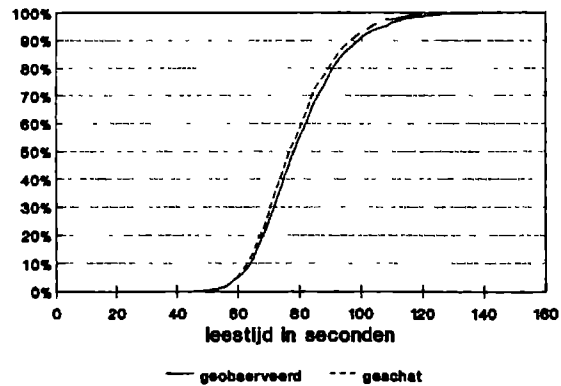
Kaart 5B



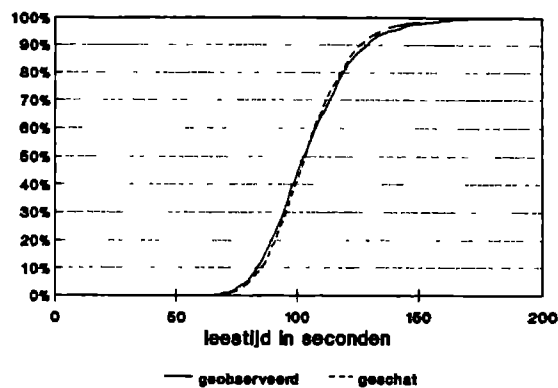
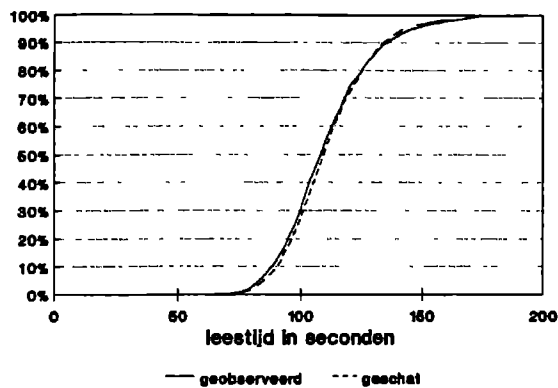
Kaart 6A



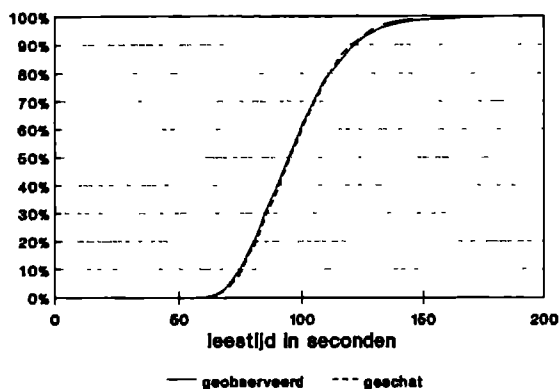
Kaart 6B



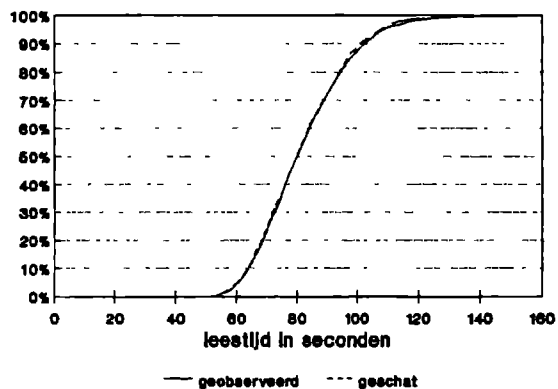
Figuur 3.4: Cumulatieve verdeling van geobserveerde en geschatte leestijden voor de



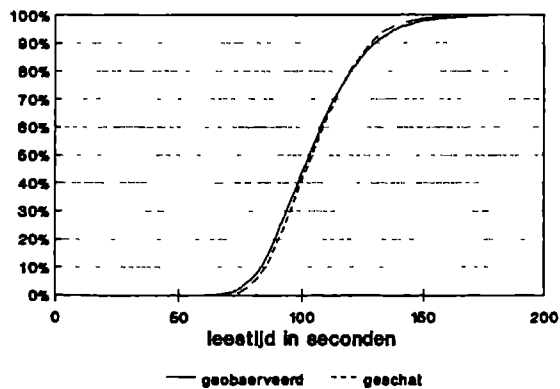
Kaart 8A



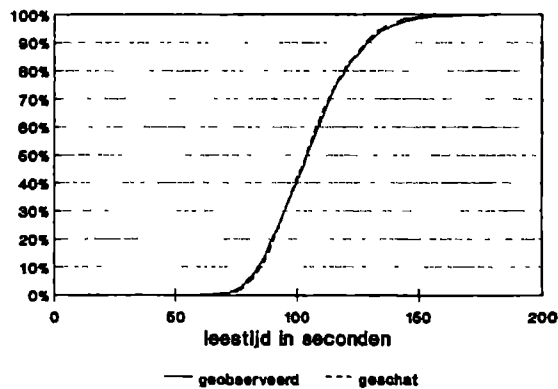
Kaart 8B



Kaart 9A



Kaart 9B



Figuur 3.5: Cumulatieve verdeling van geobserveerde en geschatte leestijden voor de

In de figuren 3.3 tot en met 3.5 zijn per toetskaart de cumulatieve verdelingen weergegeven van de geobserveerde en geschatte leestijden voor het totale gewogen bestand. Daarbij zijn alleen voor die leerlingen de leestijden op een toetskaart geschat die de betreffende kaart ook daadwerkelijk hebben gelezen. Voor iedere kaart zijn de gegevens van leerlingen met extreme geobserveerde of geschatte leestijden niet in de figuren opgenomen. De extreme leestijden zijn die tijden die meer dan drie interkwartiel afstanden verwijderd zijn van de grenzen van een boxplot (het 25ste en 75ste percentiel).

Uit de figuren kan worden afgelezen dat de cumulatieve verdelingen van de geobserveerde en geschatte leestijden redelijk dicht bij elkaar liggen. Wel zien we bij de meeste kaarten hetzelfde systematische patroon van afwijkingen. Voor de betere lezers (met lage leestijden) worden de leestijden door het model enigszins overschat, terwijl voor de zwakkere lezers (met hoge leestijden) de leestijden worden onderschat. De oorzaak van deze lichte vertekening is (nog) niet duidelijk (zie ook Jansen, 1997). De enige duidelijke uitzondering op dit patroon is kaart 6B waar de cumulatieve verdeling van de geschatte leestijden vrijwel constant boven die van de geobserveerde leestijden blijft. Dit zou betekenen dat bij deze kaart voor vrijwel alle leerlingen de leestijd enigszins onderschat wordt.

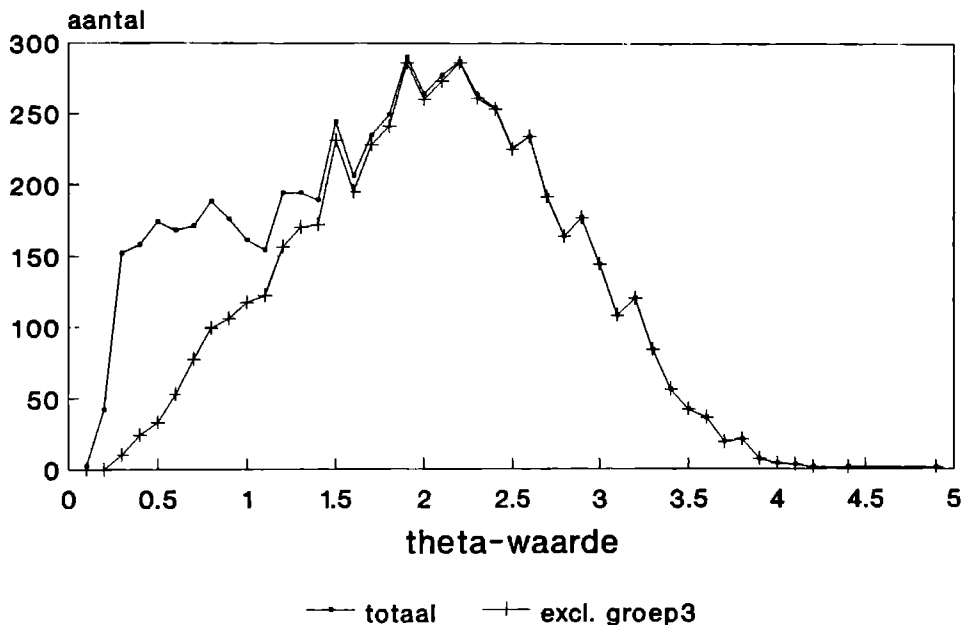
3.2.3 Grenswaarden op basis van de leestijden

De volgende stap die gezet moest worden was het opdelen van de continue leesvaardigheidsschaal $\hat{\theta}$ in de negen AVI-niveaus. De verdeling van de geschatte leesvaardigheid $\hat{\theta}$ voor de gehele periode van drie leerjaren (gebaseerd op het totale gewogen bestand) is afgebeeld in figuur 3.6. De verdeling is tweetoppig. De eerste top wordt veroorzaakt door de gegevens van de leerlingen uit maart van groep 3. Dit tijdstip bleek namelijk voor veel leerlingen nog te vroeg voor het afnemen van de AVI-toets. Zonder deze groep lijkt de verdeling veel op een normaal-verdeling (Kolmogorov-Smirnov $Z = 1,42$; $p=0,04$).

Voor het opdelen van de theta-schaal zijn eerst de ondergrens van niveau 1 en de bovengrens van niveau 9 vastgelegd. Vervolgens is de schaal in negen gelijke stukken verdeeld.

Voor het begin- en eindpunt is gezocht naar reële leessnelheden die behoren bij AVI-niveaus. Struiksma, Van der Leij en Vieijra (1995) hanteren als vuistregel: 40-50 woorden per minuut in AVI-1, 50-60 in AVI-2 en AVI-3 tot uiteindelijk 120 woorden per minuut in AVI-9. Voor de oude toetskaarten komt de ondergrens van niveau 1 overeen met een leessnelheid van 43 (oude kaart 1A) respectievelijk 36 woorden (oude kaart 1B) de bovengrens voor AVI-niveau 9 met 162 respectievelijk 144 woorden voor de oude kaarten 9A en 9B.

We hebben vervolgens voor de nieuwe kaarten van niveau 1 deze leessnelheden (tussen 36 en 43 woorden per minuut) voor de ondergrens omgezet in de leestijd (in seconden) voor de gehele tekst. Met behulp van formule 3.4 hebben we deze leestijden vervolgens omgezet in bijbehorende theta-waarden. De leestijden bleken overeen te komen met een thetawarde tussen 0,50 en 0,55 voor de ondergrens van niveau 1. Voor de bovengrens van niveau 9 kwamen we op een analoge wijze uit op thetawarden tussen 2,85 en 2,90, uitgaande van een leessnelheid tussen de 120 en 144 woorden per minuut.



Figuur 3.6: Vergelijking verdeling van de geschatte leesvaardigheid $\hat{\theta}$ met en zonder de gegevens van de leerlingen van maart groep 3

De thetawwaarden tussen 0,50 en 0,55 en de waarden tussen 2,85 en 2,90 hebben we als uitgangspunt voor het zoeken naar grenswaarden genomen. Om de rest van onze werkwijze te illustreren hebben we in tabel 3.5 voor een aantal thetawwaarden de bijbehorende geschatte leestijd in seconden voor ieder van de 18 kaarten opgenomen, zoals die met behulp van formule 3.4 berekend kan worden.

In tabel 3.5 is de thetaschaal in negen gelijke delen verdeeld met als beginpunt 0,55 en eindpunt 2,85. Elk deel, dat een interval beslaat op de thetaschaal van 0,256 zoals in de bovenste regels van de tabel te zien is, staat voor een AVI-niveau. De bovengrens van het ene niveau is tegelijkertijd de ondergrens van een niveau hoger. Door het toepassen van formule (3.4) zijn direct de grenzen voor de leestijden in seconden beschikbaar. De dubbelomrande cellen bevatten de leestijden die horen bij de betreffende grens-theta's. Zo ligt de ondergrens (voor de overgang van onvoldoende naar voldoende) voor niveau 1 bij de theta-waarde 0,55, hetgeen overeenkomt met een leestijd van 130 seconden op kaart 1A en 146 seconden op kaart 1B. Leerlingen die langer lezen over die kaarten, krijgen de score 'onvoldoende' op die kaarten en beheersen AVI-niveau 1 nog niet. Als bovengrens (voor de overgang van voldoende naar goed) voor AVI-1 geldt een theta-waarde van 0,81, hetgeen overeenkomt met een leestijd van 89 seconden bij kaart 1A en van 100 seconden bij kaart 1B. Leerlingen die dit criterium halen, krijgen de score 'goed' op de kaart.

Tabel 3.5: Geschatte leestijd (in seconden) per toetskaart voor een aantal thetawwaarden

	grenswaarden op thetaschaal voor AVI-niveau:									
	1	2	3	4	5	6	7	8	9	
kaart	0,55	0,81	1,06	1,32	1,57	1,83	2,08	2,34	2,59	2,85
1a	130	89	67	54	45	39	34	31	28	25
1b	146	100	76	61	51	44	39	34	31	28
2a	186	127	96	78	65	56	49	44	39	36
2b	182	124	94	76	64	55	48	43	39	35
3a	247	168	128	103	86	74	65	58	52	48
3b	259	177	134	108	91	78	68	61	55	50
4a	287	196	149	120	100	86	76	67	61	55
4b	257	176	133	107	90	77	68	60	55	50
5a	310	212	161	130	108	93	82	73	66	60
5b	305	208	158	127	107	92	81	72	65	59
6a	283	193	147	118	99	85	75	67	60	55
6b	288	197	149	120	101	87	76	68	61	56
7a	444	303	230	185	155	133	117	104	94	86
7b	422	288	219	176	148	127	111	99	89	81
8a	451	308	234	188	158	136	119	106	96	87
8b	383	262	199	160	134	115	101	90	81	74
9a	520	355	270	217	182	157	137	122	110	100
9b	522	357	271	218	183	157	138	123	111	101

De eerder genoemde ondergrenzen voor niveau 1 (0,50-0,55) en bovengrenzen (2,85-2,90) voor niveau 9 leveren vier mogelijke combinaties op van onder- en bovengrenzen. Ten behoeve van de hanteerbaarheid van de grenswaarden in de praktijk hebben we er voor gekozen de leestijden af te ronden op veelvouden van 5 seconden. Voor een overzicht van de uitgeteste grenswaarden verwijzen we naar Visser en Voeten (1994). De

grenswaarden zijn toegepast op het totale gewogen bestand. Prestaties van leerlingen die een bepaalde kaart hadden gelezen, konden met deze waarden worden ingedeeld in de categorieën onvoldoende, voldoende en goed op die kaart.

Op drie manieren zijn we nagaan in hoeverre de uitkomsten van de verschillende kaarten consistent waren, namelijk door:

- 1 het vergelijken van de uitkomsten van twee kaarten van hetzelfde niveau;
- 2 het vergelijken van de uitkomsten van twee opeenvolgende kaarten van eenzelfde versie;
- 3 het vergelijken van de indeling in AVI-niveaus op basis van de twee versies van de toets.

Het eerste punt betreft de overeenkomst op kaartniveau tussen de twee versies: de kans op verschillende resultaten voor twee kaarten van hetzelfde niveau moet zo klein mogelijk zijn (eerste criterium uit paragraaf 3.2.1). Dit betekent dat in de resulterende kruistabellen zoveel mogelijk de diagonalen gevuld dienen te zijn (zie figuur 3.7).

		niveau X versie A		
		onvoldoende	voldoende	goed
niveau X versie B	onvoldoende	+	-	--
	voldoende	-	+	-
	goed	--	-	+

Figuur 3.7: Optimale invulling van de kruistabel bij vergelijken van twee kaarten van eenzelfde AVI-niveau

De uiteindelijke grenswaarden dienden we zo te kiezen dat de diagonaal zoveel mogelijk gevuld is, en de overige cellen zo min mogelijk, met name de cel linksonder en rechtsboven. Tevens diende de kruistabel zoveel mogelijk symmetrisch te zijn, ofwel: de ene versie mocht niet consistent moeilijker zijn dan de andere.

Het tweede punt, consistente uitkomsten bij twee opeenvolgende kaarten, betekent dat (criterium 2 en 3 uit paragraaf 3.2.1):

- de kans op een voldoende of goed resultaat op de hoogste van twee kaarten, gegeven een onvoldoende of voldoende resultaat op de lagere kaart, zo klein mogelijk moet zijn;
- de kans op een onvoldoende resultaat op de hoogste van twee kaarten, gegeven een goed resultaat op de lagere kaart, zo klein mogelijk moet zijn.

Figuur 3.8 geeft de ideale celvulling voor de kruistabel tussen twee kaarten van opeenvolgende niveaus weer.

		niveau X+1 versie A		
		onvoldoende	voldoende	goed
niveau X versie A	onvoldoende	+	type I	
	voldoende	+		
	goed	type II	+	+

Figuur 3.8: Optimale invulling van kruistabel tussen twee kaarten van opeenvolgend AVI-niveau

De inconsistente uitkomsten zijn onderverdeeld in type I en II en corresponderen met het tweede respectievelijk derde criterium in paragraaf 3.2.1. Type I betreft die gevallen waarin de prestatie op de hogere kaart voldoende of goed is en die op de lagere voldoende of onvoldoende. Type II betreft die gevallen waarin de prestatie op de lagere van twee kaarten goed is, maar die op de hogere onvoldoende. Gezien de afnameprocedure van de AVI-toets zijn fouten van het type I ernstiger. Immers, in de afnameprocedure (zie paragraaf 1.1.2) wordt bij een score 'goed' op een kaart telkens een kaart van een hoger niveau aangeboden totdat de score 'voldoende' op een kaart wordt bereikt. Bij een inconsistentie van type I wordt niet doorgetoetst op basis van de prestatie op de lagere kaart, terwijl de leerling feitelijk op de hogere kaart 'voldoende' of 'goed' zou presteren, en dus toch leesstof van een hoger AVI-niveau beheerst.

Op basis van deze consistentie-checks zijn de uiteindelijke grenswaarden voor de leestijden gekozen (zie tabel 3.8). De uitkomsten van deze controles op basis van het toepassen van alleen de grenswaarden voor de leestijden, presenteren we hier niet, aangezien dat te omvangrijk zou zijn. Wel zullen we in paragraaf 3.3 de resultaten van vergelijkbare controles presenteren bij het toepassen van grenswaarden voor zowel de leestijden als het aantal fouten.

3.2.4 Correctie voor leesfouten

Als uitgangspunt bij het bepalen van de maximaal toegestane aantallen fouten gold dat de leessnelheid niet ten koste mocht gaan van de nauwkeurigheid. Alvorens de grenswaarden voor de fouten te bepalen, is nagegaan wat de relatie is tussen de leestijd en het aantal gemaakte fouten. In tabel 3.6 staat per kaart de correlatiecoëfficiënt weergegeven, zowel voor het totale gewogen bestand als per tijdstip, mits minimaal 50 leerlingen de betreffende kaart hebben gelezen.

Tabel 3.6: Correlaties tussen leestijd en het aantal fouten per kaart, voor het totale bestand en per tijdstip ($N \geq 50$)

kaart	totaal	maart groep 3	oktober groep 4	maart groep 4	oktober groep 5	maart groep 5	oktober groep 6	maart groep 6
1A	0,60	0,59	0,48	0,31				
1B	0,55	0,54		0,44				
2A	0,63	0,61	0,50	0,44	0,32			
2B	0,60	0,57		0,35				
3A	0,54	0,67	0,52	0,44	0,42	0,34		
3B	0,54	0,66		0,41		0,26		
4A	0,31	0,36	0,40	0,29	0,32	0,28	0,22	
4B	0,36	0,57		0,33		0,18		
5A	0,21		0,18	0,23	0,19	0,14	0,05	0,02
5B	0,24			0,22		0,16		0,07
6A	0,22		0,34	0,22	0,21	0,20	0,21	0,15
6B	0,17			0,22		0,15		0,13
7A	0,22			0,21	0,25	0,23	0,23	0,04
7B	0,16			0,19		0,16		0,14
8A	0,38			0,33	0,37	0,31	0,39	0,40
8B	0,23			0,23		0,22		0,19
9A	0,30			0,25	0,26	0,31	0,23	0,37
9B	0,34			0,39		0,31		0,31

De correlatie tussen de tijd en het aantal fouten is het hoogst bij de toetskaarten van de eerste drie niveaus, namelijk tussen 0,54 en 0,60 (totale bestand). De correlaties voor de andere toetskaarten zijn beduidend lager: tussen 0,16 en 0,38. Het laagst zijn de correlaties bij de toetskaarten van niveau 5 tot en met 7. Vergelijking van de correlaties op de verschillende tijdstippen in de basisschoolperiode laat zien dat de samenhang in veel gevallen afneemt in de loop van de basisschoolperiode. Zo bedraagt bijvoorbeeld de correlatie tussen tijd en fouten van kaart 1A 0,59 in maart van groep 3, 0,48 in oktober van groep 4 en 0,31 in maart van groep 4.

De bevinding dat de samenhang tussen de leessnelheid en de nauwkeurigheid afneemt in de loop van de basisschoolperiode komt overeen met wat we op basis van de theorie

omtrent de ontwikkeling van de decodeervaardigheid mochten veronderstellen. Daarbij gaan we er immers vanuit dat kinderen eerst leren de woorden nauwkeurig te herkennen en daarna de woorden steeds sneller gaan decoderen (zie onder andere Ehri & Wilce, 1979, 1983). Al vrij snel verschilt de nauwkeurigheid waarmee kinderen woorden lezen niet meer, terwijl de verschillen in leessnelheid nog wel groot zijn. Dit wordt weerspiegeld door de bevinding dat de correlaties tussen nauwkeurigheid (of het aantal fouten) en de leessnelheid afnemen naarmate de tijd vordert.

De lage correlaties tussen de leestijden en het aantal fouten voor met name de kaarten van de hogere niveaus en voor leerlingen van de hogere leerjaren rechtvaardigen de keuze om zowel criteria voor leestijden als criteria voor fouten te hanteren bij het bepalen van het AVI-niveau. Immers een grotere snelheid gaat niet altijd gepaard met een grotere nauwkeurigheid.

Om voor iedere toetskaart de onder- en bovengrens te bepalen is als volgt gewerkt.

Startpunt vormde de indeling in AVI-niveaus op basis van de thetaverdeling van de leestijden.

Voor iedere toetskaart zijn het gemiddelde en de standaardafwijking voor het aantal fouten berekend per AVI-niveau volgens de eerder besproken vier verdelingen van de theta-schaal. Per kaart zijn de volgende grenswaarden uitgetest:

- 1 voor de *voldoende-grens* van een kaart van niveau X:
 - a: het gemiddelde aantal fouten + 0,5 standaardafwijking van degenen die volgens de theta-verdeling in niveau X lezen;
 - b: het gemiddelde aantal fouten + 1 standaardafwijking van degenen die volgens de theta-verdeling in niveau X lezen;
- 2 voor de *goed-grens* van een kaart van niveau X:
 - a: het gemiddelde aantal fouten + 0,5 standaardafwijking van degenen die volgens de theta-verdeling in niveau X+1 lezen;
 - b: het gemiddelde aantal fouten + 1 standaardafwijking van degenen die volgens de theta-verdeling in niveau X+1 lezen.

In bijlage 10 staan de resulterende grenswaarden voor de fouten weergegeven. Deze grenzen zijn toegepast samen met de uiteindelijke grenzen voor de leestijden volgens het principe dat de prestatie op een kaart alleen goed is, wanneer zowel de tijd als het aantal fouten aan het criterium voor 'goed' voldoen, en dat de score onvoldoende is als òf de tijd òf het aantal fouten onvoldoende is.

Op dezelfde wijze als bij het toepassen van de grenswaarden voor de leestijden alleen, is met behulp van kruistabellen de consistentie van de resultaten nagegaan. Tevens is voor iedere toetskaart nagegaan in hoeveel gevallen de uitkomst van die kaart (goed, voldoende, onvoldoende) lager uitkomt bij het toepassen van de grenswaarden voor fouten. In tabel 3.7 is voor iedere toetskaart het percentage gevallen weergegeven waarvoor geldt dat de uitkomst lager uitvalt dan bij het toepassen van enkel de grenswaarden voor de leestijden.

Tabel 3.7: Percentage gevallen waarin de uitkomst lager uitvalt door toepassing van het foutencriterium bovenop het leestijdencriterium

kaart	fouten: gemiddelde plus 0,5 standaardafw.				fouten: gemiddelde plus 1 standaardafw			
	vold. -> onvold.*	goed -> vold.*	goed -> onvold.*	totaal**	vold. -> onvold.*	goed -> vold.*	goed -> onvold.*	totaal**
1A	25	13	9	17	15	11	4	11
1B	30	6	12	15	19	7	5	9
2A	24	7	10	13	16	5	5	8
2B	24	10	13	13	16	6	6	8
3A	31	7	15	20	17	9	6	12
3B	30	6	12	17	19	8	4	11
4A	29	0	17	17	15	4	7	10
4B	24	9	15	20	15	0	9	9
5A	16	11	12	17	9	4	7	9
5B	24	10	14	20	15	5	9	12
6A	20	12	14	20	13	6	7	11
6B	35	0	25	23	19	0	13	12
7A	22	8	16	17	14	5	10	11
7B	20	7	18	17	15	0	13	10
8A	23	6	11	14	11	6	4	8
8B	25	10	16	21	18	8	8	13
9A	33	9	16	18	15	10	6	10
9B	31	10	15	19	11	9	6	10

* Percentering op basis van het aantal leerlingen dat oorspronkelijk in de eerst genoemde categorie (voldoende respectievelijk goed) voor de betreffende kaart zit

** Percentering op basis van het totaal aantal leerlingen dat de betreffende kaart heeft gelezen

De tabel kan als volgt gelezen worden. De tweede kolom bevat per kaart het percentage leerlingen dat met alleen de grenswaarde voor tijden voldoende scoort, maar door toevoeging van de grenswaarde voor fouten (0,5 standaardafwijking) onvoldoende scoort (gepercenteerd ten opzichte van alle leerlingen die oorspronkelijk voldoende scoorden op de betreffende kaart). De derde en vierde kolom geven overeenkomstige percentages voor de leerlingen die met alleen de grenswaarden voor tijden goed scoorden maar door de

toevoeging van de grenswaarden voor fouten respectievelijk voldoende en onvoldoende scoren (gepercenteerd ten opzichte van alle leerlingen die oorspronkelijk goed scoorden op de bijbehorende kaart). De vijfde kolom bevat het totale percentage leerlingen dat op een lager niveau uitkomt door toevoeging van het foutencriterium (gepercenteerd ten opzichte van het totaal aantal leerlingen dat de betreffende kaart heeft gelezen). De zesde tot en met de negende kolom geven dezelfde gegevens bij toepassing van een soepeler foutencriterium. De hoogste percentages vinden we in de tweede en zesde kolom, hetgeen betekent dat de meeste verschuivingen in niveau plaatsvinden door een terugval van voldoende naar onvoldoende.

Als basis voor de uiteindelijke grenzen hebben we gekozen voor het gemiddelde plus 1 standaardafwijking. Deze grenswaarden zijn wel enigszins bijgesteld. Eerst is er voor gezorgd dat de foutencriteria voor voldoende en goed verschillend zijn. In een aantal gevallen waren deze namelijk gelijk (zie bijlage 10). Daarnaast is geprobeerd om het toegestane percentage fouten voor twee kaarten van hetzelfde niveau gelijk te houden. Daarbij is uiteraard gekeken naar de mate van consistentie van de uiteindelijke uitkomsten tussen kaarten van een verschillende versie en tussen elkaar opvolgende kaarten van dezelfde versie.

3.2.5 De uiteindelijke grenswaarden

Tabel 3.8 bevat de uiteindelijke grenswaarden die we hebben gekozen. Bij de leestijden staan tevens de bijbehorende theta-waarde en de leesnelheid in aantal gelezen woorden per minuut vermeld, bij het aantal fouten tevens het percentage fouten ten opzichte van het totale aantal woorden.

De tweede en zesde kolom geven de theta-waarden weer die horen bij de leestijden voor de overgang van onvoldoende naar voldoende, respectievelijk van voldoende naar goed. In paragraaf 3.2.3 hebben we als uitgangspunt gekozen dat de afstanden tussen de grenswaarden op de thetaschaal voor de leestijden gelijk moeten zijn. Wanneer we nu de thetawwaarden behorend bij de uiteindelijke grenswaarden voor de leestijden bekijken, zien we die gelijkheid in afstanden niet altijd terug. Bijvoorbeeld de thetawwaarden behorende bij de voldoende-grenzen voor de kaarten 2B, 3B en 4B bedragen achtereenvolgens: 0,77, 1,02, 1,23. De afstanden tussen deze thetawwaarden zijn niet gelijk als gevolg van het feit dat de leestijden afgerond zijn op veelvoud van vijf seconden. Daardoor zijn ook de thetawwaarden behorend bij de grenswaarden van twee kaarten van eenzelfde niveau niet geheel gelijk aan elkaar. Zo zien we bijvoorbeeld voor de overgang van voldoende naar goed de waarde 0,75 voor kaart 1A en de waarde 0,77 voor kaart 1B. Deze afronding is een bron van inconsistentie van resultaten op de twee toetsversies. Bovendien kan deze afronding bijdragen aan inconsistenties in resultaten op opeenvolgende kaarten.

Tabel 3.8: Uiteindelijke grenswaarden per kaart voor de leestijden en het aantal fouten

kaart	grenswaarden voor de leestijd						grenswaarden voor aantal fouten			
	voldoende			goed			voldoende		goed	
	sec	theta	wrd/ min	sec	theta	wrd/ min	aantal	pct	aantal	pct
1A	135	0,53	46,2	95	0,75	65,7	6	5,8	4	3,8
1B	150	0,54	41,6	105	0,77	59,4	6	5,8	4	3,8
2A	135	0,76	57,3	100	1,02	77,4	5	3,9	4	3,1
2B	130	0,77	57,2	95	1,05	78,3	5	4,0	4	3,2
3A	135	1,00	58,7	105	1,29	75,4	7	5,3	5	3,8
3B	140	1,02	61,3	110	1,30	78,0	8	5,6	6	4,2
4A	125	1,26	71,5	105	1,50	85,1	6	4,0	5	3,4
4B	115	1,23	81,4	95	1,49	98,5	6	3,8	5	3,2
5A	110	1,55	89,5	95	1,80	103,6	5	3,0	4	2,4
5B	110	1,53	84,0	95	1,77	97,3	5	3,2	4	2,6
6A	90	1,73	100,0	80	1,95	112,5	4	2,7	3	2,0
6B	95	1,67	95,4	85	1,86	106,6	4	2,6	3	2,0
7A	120	2,03	103,0	105	2,32	117,7	6	2,9	5	2,4
7B	115	2,02	103,8	100	2,32	119,4	6	3,0	5	2,5
8A	110	2,25	111,3	100	2,48	122,4	7	3,4	5	2,5
8B	95	2,22	114,9	85	2,48	128,5	7	3,8	5	2,7
9A	110	2,60	125,5	100	2,86	138,0	5	2,2	4	1,7
9B	115	2,50	117,9	105	2,74	129,1	5	2,2	4	1,8

3.3 De betrouwbaarheid van de AVI-toetskaarten

In deze paragraaf gaan we na hoe consistent de uitkomsten van de AVI-toetskaarten zijn (onderzoeksvraag 2), waarbij we achtereenvolgens de drie criteria, geformuleerd in paragraaf 3.2.1, zullen bespreken.

Het eerste criterium betreft de overeenkomst in resultaat van twee kaarten van hetzelfde AVI-niveau. De kans op een verschillend resultaat voor twee kaarten van hetzelfde niveau

zou zo klein mogelijk moeten zijn. In tabel 3.9 staat het percentage inconsistente en consistente uitkomsten weergegeven bij vergelijking van de resultaten van twee kaarten van hetzelfde niveau.

Tabel 3.9: Percentage consistente en inconsistente uitkomsten bij vergelijking van de resultaten op de kaarten van eenzelfde niveau van versie A en B

niveau	inconsistent		consistent	N
	A lager	B lager		
1	12,5	13,0	74,5	1124
2	8,4	9,2	82,4	1277
3	10,1	15,1	74,8	810
4	14,6	16,2	69,2	799
5	18,4	19,2	62,4	781
6	19,6	16,0	64,4	830
7	14,3	21,6	64,1	945
8	15,4	14,8	69,8	1388
9	18,7	12,2	69,1	1064

De vierde kolom geeft het percentage leerlingen aan waarvoor geldt dat de resultaten op twee verschillende kaarten van hetzelfde AVI-niveau overeenkomen. Het percentage consistente uitkomsten voor twee kaarten van hetzelfde niveau ligt tussen de 62% en 82%. De tweede en derde kolom bevatten het percentage inconsistente uitspraken uitgesplitst naar twee mogelijkheden: de uitkomst van de kaart van niveau A is lager dan dat van niveau B, respectievelijk het omgekeerde. Voor elk paar kaarten van hetzelfde niveau is getoetst in hoeverre sprake is van symmetrie (vergelijk figuur 3.7). Voor drie van de negen niveaus bleek het symmetrie-model verworpen te moeten worden: voor niveau 4, 7 en 9. Voor kaart 4 is deze asymmetrie niet direct uit tabel 3.9 af te lezen. Over het geheel genomen is het percentage lagere uitkomsten voor versie 4A ongeveer gelijk aan dat voor versie 4B. Echter een voldoende prestatie op 4A bij een goede prestatie op 4B komt minder vaak voor dan het omgekeerde. Tevens komt een onvoldoende prestatie op 4B bij een goede prestatie op 4A vaker voor dan het omgekeerde.

Uit deze resultaten kunnen we afleiden dat de kans op gelijke uitkomsten van kaarten van hetzelfde AVI-niveau tussen de 62% en 82% ligt. Het percentage consistente uitkomsten is het hoogst voor de niveaus 1, 2 en 3 (75%-82%) en de niveaus 4, 8 en 9 (69%-70%). Met name voor de niveaus 5, 6 en 7 is de kans op inconsistente uitkomsten groot. Voor de meeste niveaus geldt echter niet dat de A- of B-versie consistent hogere of lagere uitkomsten geeft. Dit alles betekent dat de vergelijkbaarheid van de uitkomsten tussen versie A en B op kaartniveau matig is. Voor een deel zullen de inconsistenties

veroorzaakt worden door het feit dat de grenzen voor de leestijden zijn afgerond tot een veelvoud van vijf seconden. Maar deze afrondingen zijn geen voldoende verklaring voor de optredende inconsistenties. We zien in tabel 3.8 namelijk dat bij de kaarten van niveau 6 en 9 de effecten van de afrondingen het grootst zijn, gezien de verschillen in thetawwaarden behorende bij de grenzen voor leestijden. Echter voor niveau 5 en 7 verschillen deze thetawwaarden nauwelijks, terwijl daar de kans op inconsistente uitkomsten het grootst is.

De overige twee criteria voor consistentie in paragraaf 3.2.1 hebben betrekking op het vergelijken van kaarten van eenzelfde versie die elkaar opvolgen qua AVI-niveau. In paragraaf 3.2.3 hebben we de bijbehorende inconsistente uitkomsten onderverdeeld in type I en II. Type I betreft die gevallen waarin de prestatie op de hogere kaart voldoende of goed is terwijl die op de lagere voldoende of onvoldoende is. Type II betreft die gevallen waarin de prestatie op de kaart van het lagere niveau goed is, maar die op de kaart van het hogere niveau onvoldoende (zie ook figuur 3.8). In tabel 3.10 staat weergegeven hoe consistent de resultaten op telkens twee onmiddellijk opeenvolgende kaarten zijn.

Deze tabel is opgebouwd overeenkomstig de twee criteria voor consistentie betreffende twee opeenvolgende kaarten. Kolom 2 en 4 zijn een nadere uitsplitsing van de inconsistenties van type I. De tweede kolom bevat de kans op een voldoende of goed resultaat op de hogere kaart, gegeven een onvoldoende resultaat op de lagere kaart. Deze kans ligt tussen de 0,03 en de 0,29. Voor een aantal paren van kaarten is deze kans groter dan 0,20: 4A-5A, 5B-6B en 7B-8B. Voor deze paren van kaarten geldt dat meer dan 20% van de leerlingen die op de lagere kaart een onvoldoende resultaat behalen, het niveau van de hogere kaart wel beheersen. Dit houdt feitelijk een verschil in uitkomst van minimaal twee niveaus in. Immers, onvoldoende op kaart 4A betekent maximaal AVI-niveau 3, terwijl voldoende of goed op 5A minimaal niveau 5 impliceert.

De vierde kolom geeft de kans weer op een voldoende of goed resultaat op de hogere kaart, gegeven een voldoende resultaat op de lagere; een verschil van minimaal één niveau. Deze kansen liggen tussen 0,18 en 0,52. Voor een aantal paren van kaarten is deze kans erg groot (meer dan 0,40): 5A-6A, 7A-8A, 3B-4B, 5B-6B en 7B-8B.

De zesde kolom geeft de kans weer op een onvoldoende resultaat op de hogere kaart gegeven een goed resultaat op de lagere (type II inconsistentie). Deze kans is voor een aantal paren groter dan 0,20: 6A-7A, 8A-9A en 6B-7B.

In de achtste kolom staat het totale percentage consistente uitkomsten weergegeven. Het percentage consistenties is het laagst voor de vergelijking van 6B met 7B (68%). Voor de overige paren is het percentage consistente uitkomsten 75% of hoger. De consistentie is het hoogst voor de kaarten van de drie laagste niveaus.

Tabel 3.10: Conditionele kansen op inconsistente uitkomsten en het percentage consistente uitkomsten bij twee onmiddellijk opeenvolgende kaarten

	kans op voldoende/goed op hogere kaart als op lagere kaart het resultaat is (type I):				kans op onvoldoende op hogere kaart als resultaat op lagere kaart goed is (type II)		consistente uitkomsten	
	onvoldoende		voldoende					
	p	N'	p	N'	p	N'	%	N''
1A-2A	0,05	444	0,24	376	0,11	743	87,3	1563
2A-3A	0,14	135	0,25	219	0,14	885	83,9	1238
3A-4A	0,13	132	0,33	207	0,16	790	81,3	1129
4A-5A	0,24	147	0,33	163	0,16	658	79,7	968
5A-6A	0,15	176	0,48	221	0,14	602	78,3	998
6A-7A	0,14	180	0,28	230	0,27	702	74,9	1111
7A-8A	0,16	289	0,41	377	0,19	560	75,2	1226
8A-9A	0,06	283	0,18	335	0,23	1239	80,5	1857
1B-2B	0,03	375	0,20	248	0,13	418	89,2	1041
2B-3B	0,11	71	0,22	102	0,17	386	82,9	559
3B-4B	0,17	65	0,43	110	0,10	384	82,9	558
4B-5B	0,11	57	0,36	120	0,17	374	79,6	550
5B-6B	0,29	95	0,52	114	0,10	329	77,8	537
6B-7B	0,19	74	0,23	114	0,36	418	68,4	605
7B-8B	0,22	201	0,50	198	0,10	307	75,1	705
8B-9B	0,07	148	0,37	212	0,18	743	79,8	1103

'N = aantal leerlingen dat op de lagere kaart de betreffende score heeft behaald

''N = totaal aantal leerlingen dat beide kaarten heeft gelezen

In paragraaf 3.2.3 hebben we al aangegeven dat inconsistenties van het type I ernstiger zijn dan die van type II. Dit omdat bij type I onterecht besloten wordt geen kaart van een hoger niveau aan te bieden. Bij drie paren van opeenvolgende kaarten is de kans op minimaal voldoende op de hogere kaart, gegeven een onvoldoende resultaat op de lagere kaart groter dan 0,20: 4A-5A, 5B-6B en 7B-8B. Voor een vijftal paren van opeenvolgende kaarten is de kans op minimaal voldoende op de hogere kaart, gegeven een voldoende resultaat op de lagere, groter dan 0,40: 5A-6A, 7A-8A, 3B-4B, 5B-6B en 7B-8B. Wanneer

we de grootte van de moeilijkheidsparameters van deze toetskaarten bekijken (tabel 3.2) dan valt op dat deze niet zo mooi verlopen. Zo is kaart 3B iets moeilijker dan 4B, is 4A iets moeilijker dan 5A, en is 5B iets moeilijker dan 6B. De kaarten 5A en 6A, 7A en 8A, en 7B en 8B zijn telkens ongeveer even moeilijk. Het lijkt er op dat het niet zo mooi verlopen van de moeilijkheidsparameters bijdraagt aan de mate van inconsistentie van uitkomsten van opeenvolgende kaarten. Voor één paar van kaarten waarvoor geldt dat de verschillen in moeilijkheidsgraad klein zijn (8A-9A), zien we echter niet zo'n grote kans op inconsistenties van type I.

Terug naar de in paragraaf 3.2.1 geformuleerde criteria kunnen we voor twee onmiddellijk opeenvolgende kaarten het volgende concluderen. De kans op inconsistenties van het type I (tweede criterium) is voor nogal wat paren van opeenvolgende kaarten groot. Dit geldt met name voor de kans op een voldoende of goed resultaat op de hogere kaart, gegeven een *voldoende resultaat* op de lagere. Deze kans is voor alle paren op één na groter dan 0,20. In een aantal gevallen is de kans zelfs zeer groot te noemen (groter dan 0,40).

Tot slot geldt voor vrijwel alle paren dat de kans op een onvoldoende resultaat op de hogere, gegeven een goed resultaat op de lagere kaart, kleiner is dan 0,20 (criterium 3). Deze type-II-inconsistentie is overigens niet zo ernstig gezien de procedure die gevolgd wordt voor het bepalen van het AVI-niveau, aangezien dit slechts betekent dat ten onrechte nog een kaart van een niveau hoger wordt aangeboden.

We kunnen concluderen dat de consistentie van de uitkomsten tussen opeenvolgende kaarten te wensen over laat. De grootste inconsistenties vinden we bij kaarten waarvan de moeilijkheidsgraden dicht bij elkaar of in omgekeerde volgorde liggen. Ook hier zal het afronden van de leestijden naar veelvoud van vijf seconden de inconsistente uitkomsten hebben versterkt. Wel is het zo dat de meeste inconsistenties wijzen op een verschil van één AVI-niveau. Om hierin nog wat meer inzicht te krijgen hebben we een tabel gemaakt analoog aan tabel 3.10, maar nu voor twee kaarten met een niveau ertussen, bijvoorbeeld 1A vergeleken met 3A. Dit is mogelijk omdat de meeste leerlingen drie opeenvolgende toetskaarten hebben gelezen. De enige inconsistenties die bij dergelijke vergelijkingen op kunnen treden hebben betrekking op type I.

De kans op een voldoende of goed resultaat op de hogere kaart, gegeven een onvoldoende resultaat op de lagere kaart, is voor een aantal paren nog vrij hoog. Voor 4A-6A, 2B-4B en 6B-8B ligt deze kans boven de 0,20. Dit impliceert feitelijk een verschil van minimaal 3 niveaus in uitkomst. Immers, onvoldoende op 4A betekent maximaal niveau 3, terwijl voldoende of goed op 6A minimaal niveau 6 impliceert.

De kans op een voldoende of goed resultaat op de hogere kaart, gegeven voldoende op de lagere is bij twee paren groter dan 0,20. Dit betekent een verschil van minimaal 2 niveaus.

Tabel 3.11: Conditionele kansen op inconsistente uitkomsten bij twee opeenvolgende kaarten met één niveau ertussen

	kans op voldoende/goed op hogere kaart als op lagere kaart het resultaat is:			
	onvoldoende		voldoende	
	p	N	p	N
1A-3A	0,03	33	0,17	86
2A-4A	0,14	36	0,12	70
3A-5A	0,13	23	0,28	56
4A-6A	0,20	61	0,14	65
5A-7A	0,06	54	0,13	102
6A-8A	0,11	57	0,11	99
7A-9A	0,03	105	0,06	173
1B-3B	0,07	21	0,11	42
2B-4B	0,20	13	0,19	33
3B-5B	0,18	14	0,25	38
4B-6B	0,06	17	0,11	45
5B-7B	0,13	28	0,15	47
6B-8B	0,31	30	0,16	44
6B-9B	0,04	69	0,10	108

N = aantal leerlingen dat op de lagere kaart de betreffende score heeft behaald

We kunnen op basis van alle voorgaande resultaten concluderen dat de consistentie in uitkomsten tussen twee toetskaarten niet bevredigend is. In de meeste gevallen impliceert een inconsistentie uiteindelijk een verschil van één AVI-niveau, maar ook grotere verschillen hebben bij bepaalde paren van kaarten een niet te verwaarlozen kans van voorkomen.

Tot slot zijn we nagegaan hoe consistent de resultaten van beide versies van de volledige AVI-toets zijn. In tabel 3.12 is de kruistabel weergegeven betreffende de relatie tussen versie A en versie B.

Tabel 3.12: Kruistabel tussen versie A en B (aantal, rijpercentage, kolompercentage en totale percentage)

AVI versie A	AVI versie B											To-taal
	< 1	1	2	3	4	5	6	7	8	9	> 9	
<1 N	376	72	7	5	1							459
rij %	81,9	15,6	1,4	1,0	0,1							12,8
kol %	80,9	22,7	2,5	1,7	0,2							
tot %	10,5	2,0	0,2	0,1	0,0							
1 N	80	180	52	15	2	1	1					330
rij %	24,2	54,5	15,8	4,4	0,6	0,3	0,2					9,2
kol %	17,2	57,2	20,0	5,4	0,7	0,4	0,1					
tot %	2,2	5,0	1,5	0,4	0,1	0,0	0,0					
2 N	8	48	132	42	3	2		1				235
rij %	3,2	20,5	56,1	17,9	1,3	0,9		0,2				6,6
kol %	1,6	15,3	50,5	15,5	1,0	0,8		0,1				
tot %	0,2	1,3	3,7	1,2	0,1	0,1		0,0				
3 N	1	13	58	147	49	20	4	2	1			294
rij %	0,3	4,3	19,6	50,1	16,7	6,8	1,4	0,5	0,3			8,2
kol %	0,2	4,0	22,1	54,3	17,1	7,8	1,0	0,4	0,3			
tot %	0,0	0,3	1,6	4,1	1,4	0,6	0,1	0,0	0,0			
4 N		1	9	39	129	39	15	3	1			235
rij %		0,4	3,8	16,6	54,9	16,6	6,2	1,1	0,4			6,6
kol %		0,3	3,5	14,4	45,1	15,1	3,7	0,7	0,3			
tot %		0,0	0,3	1,1	3,6	1,1	0,4	0,1	0,0			
5 N		2	4	19	75	122	80	16	1	1		318
rij %		0,5	1,1	6,0	23,6	38,4	25,0	5,0	0,2	0,3		8,9
kol %		0,5	1,3	7,0	26,2	47,4	20,0	4,4	0,1	0,5		
tot %		0,0	0,1	0,5	2,1	3,4	2,2	0,4	0,0	0,0		
6 N			1	4	21	51	171	66	16	5	2	335
rij %			0,1	1,2	6,3	15,2	51,0	19,6	4,8	1,3	0,4	9,3
kol %			0,2	1,5	7,3	19,8	42,9	18,1	4,8	2,3	0,3	
tot %			0,0	0,1	0,6	1,4	4,8	1,8	0,4	0,1	0,0	
7 N				1	4	18	98	191	65	16	7	399
rij %				0,1	0,9	4,5	24,5	47,9	16,3	4,0	1,8	11,1
kol %				0,2	1,2	7,0	24,6	52,8	19,3	8,3	1,6	
tot %				0,0	0,1	0,5	2,7	5,3	1,8	0,4	0,2	
8 N					3	4	27	65	161	49	41	349
rij %					0,9	1,0	7,7	18,7	46,1	14,1	11,6	9,7
kol %					1,0	1,4	6,8	18,0	47,8	25,5	9,2	
tot %					0,1	0,1	0,8	1,8	4,5	1,4	1,1	
9 N						1	3	16	59	76	65	218
rij %						0,5	1,1	7,1	26,8	34,9	29,6	6,1
kol %						0,4	0,6	4,3	17,4	39,5	14,7	
tot %						0,0	0,1	0,4	1,6	2,1	1,8	
>9 N							1	4	34	46	326	410
rij %							0,2	1,0	8,2	11,2	79,4	11,5
kol %							0,3	1,1	10,0	23,9	74,1	
tot %							0,0	0,1	0,9	1,3	9,1	
Tot N	465	315	261	271	286	258	397	362	336	193	439	3580
%	13,0	8,8	7,3	7,6	8,0	7,2	11,1	10,1	9,4	5,4	12,3	100

Het percentage overeenkomstige resultaten (zie omrande diagonaal) tussen versie A en B bedraagt 56,1%, hetgeen vrij laag is. In 21,2% van de gevallen is de uitkomst van A lager dan die van B, en voor 22,7% geldt het omgekeerde. In die gevallen dat de uitkomsten van versie A en B niet overeenkomen, betreft het meestal één niveau verschil.

Het tegenvallende percentage overeenkomende resultaten kan uiteraard deels worden verklaard door de al eerder gevonden inconsistenties tussen de resultaten van de afzonderlijke kaarten. We hebben echter ook nog gezocht naar additionele verklaringen. Een eerste mogelijkheid die we hebben onderzocht, is gelegen in de manier waarop in dit onderzoek de toets is afgenomen. In de regel is de procedure van het afnemen van de AVI-toetskaarten zo dat een leerling bij de score 'goed' telkens een kaart van één niveau hoger leest, totdat de score op een kaart 'voldoende' is (zie paragraaf 1.1.2). Het AVI-niveau is dan het niveau van de kaart waarop de score 'voldoende' is behaald. In het onderzoek is echter een andere procedure gevolgd bij het aanbieden van kaarten aangezien de grenswaarden om vast te stellen of de prestatie 'goed', 'voldoende' of 'onvoldoende' is met dit onderzoek bepaald moeten worden. Op basis van een inschatting van het leesniveau is bepaald welke drie kaarten een leerling te lezen kreeg (zie paragraaf 3.1.2). Als de inschatting niet geheel juist is, kunnen sommige leerlingen alleen kaarten van een te hoog niveau of alleen kaarten van een te laag niveau hebben gelezen. In dit onderzoek is in het eerste geval het AVI-niveau gezet op het niveau van de laagste kaart minus 1, en in het tweede geval op het niveau van de hoogste kaart plus 1. Het blijkt dat 6,4% respectievelijk 5,5% van de leerlingen alleen kaarten van een te hoog niveau van versie A respectievelijk versie B heeft gelezen. Daarnaast heeft 12,4% respectievelijk 11,7% alleen kaarten van een te laag niveau gelezen van versie A respectievelijk versie B. Wanneer we nu alleen voor de leerlingen die wel de goede kaarten hebben gelezen, de uitkomsten van versie A en versie B vergelijken (gewogen $N=2716$), dan is het percentage overeenkomstige resultaten iets hoger: 62,2%. Dit betekent een kleine toename in overeenstemming van zo'n 6%, waaruit we moeten afleiden dat de manier waarop de toets is afgenomen, geen voldoende verklaring is voor de lage overeenstemming.

De tweede mogelijkheid die is onderzocht betreft de procedure voor het bepalen van het AVI-niveau. Deze procedure is vrij streng. Alleen wanneer een leerling op een kaart de score 'goed' behaalt, krijgt hij een hogere kaart aangeboden. In een aantal gevallen blijkt een leerling met de score 'voldoende' op een kaart ook een score 'voldoende' op de kaart van een niveau hoger te behalen. Een andere, wat soepelere procedure, zou zijn dat men doortoetst tot de leerling de score 'onvoldoende' behaalt op een kaart. Het AVI-niveau is dan het niveau van de vorige kaart, waarop de leerling nog wel 'voldoende' scoort. De goed-grenzen worden in dat geval niet gebruikt. Deze procedure is eveneens toegepast op de gegevens van dit onderzoek. Voor nogal wat leerlingen geldt echter dat niet lang genoeg doorgetoetst is voor deze soepelere procedure, waarbij immers minimaal één kaart meer aangeboden wordt dan bij de normale procedure: 27,7% respectievelijk 26,9% van de leerlingen (voor een deel uiteraard dezelfde leerlingen) had feitelijk nog een kaart van een niveau hoger van versie A respectievelijk versie B moeten lezen. De gegevens van deze leerlingen zijn niet betrokken in de analyses.

Door toepassing van de soepelere procedure is het percentage overeenkomstige resultaten toegenomen tot 67,9% (gewogen $N=2100$). Vergelijking van de resulterende AVI-niveaus met de oorspronkelijke AVI-niveaus laat zien dat met de soepelere procedure 7,5% (versie A) respectievelijk 8,6% (versie B) van de leerlingen een hoger AVI-niveau behaalt. Dit betreft vrijwel altijd één niveau hoger.

De vergelijking van de oude en de nieuwe AVI-toets

Met behulp van de gegevens uit de tweede try-out (zie paragraaf 2.2.4) kunnen we de overeenkomst nagaan tussen de oude en de nieuwe AVI-toets. Bij deze leerlingen zijn immers gegevens bekend zowel over het niveau bepaald met de oude AVI-toets (versie A) als over het niveau bepaald met de nieuwe AVI-toets. Op voorhand verwachten we geen al te hoge overeenkomst in resultaten. Een belangrijk kritiekpunt op de oude AVI betrof namelijk de verdeling van de leerlingen over de AVI-niveaus. Deze was niet gelijkmatig te noemen. Een veel gehoorde opmerking uit de praktijk was dat leerlingen op het ene AVI-niveau leken te blijven hangen, terwijl ze door andere niveaus heel snel heen 'schooten'. Bij de nieuwe AVI-niveaus hebben we getracht die verdeling gelijkmatiger te laten zijn.

Voor de analyses hebben we die leerlingen geselecteerd die:

- 1 niet alleen te hoge of te lage kaarten van de nieuwe AVI-toets hebben gelezen, en
- 2 volgens de normen van Dickhout-Rutten (1983) ingedeeld konden worden met de oude AVI-toets (versie A).

Dit laatste behoeft enige toelichting. In de try-out zijn de toetsen afgenomen door de scholen zelf en zijn de leerlingen ook door de scholen zelf ingedeeld naar AVI-niveau met de oude AVI-toets. Het bleek echter dat dat laatste niet altijd juist gebeurde, waardoor het niveau waarin een leerling werd ingedeeld niet altijd overeenkwam met het niveau dat het zou moeten zijn volgens de leestijd en het aantal fouten op de oude toets.

Tabel 3.13 geeft het verband weer tussen de niveaus volgens de oude (versie A) en de nieuwe toets (versie A en B) voor de resterende leerlingen. De tabel kan als volgt gelezen worden. In de eerste kolom staat het niveau bepaald met behulp van versie A van de oude AVI-toets. De tweede tot en met de vierde kolom bevatten per AVI-niveau volgens de oude toets, het percentage leerlingen dat met versie A van de nieuwe toets op hetzelfde, een lager of een hoger niveau is ingedeeld. De kolommen zes tot en met acht bevatten dezelfde gegevens maar nu voor versie B van de nieuwe toets.

Het percentage overeenstemming bedraagt 41,0% respectievelijk 41,3% tussen de oude toets en versie A respectievelijk versie B van de nieuwe toets, hetgeen vrij laag is. Het percentage overeenstemming verschilt wel duidelijk afhankelijk van het niveau op de oude toets. Over het algemeen zijn de niveaus op de oude toets hoger dan op de nieuwe, waarbij het vrijwel altijd een verschil van één niveau betreft. Uitzondering is niveau 9, waar ruim de helft van de leerlingen met de nieuwe AVI-toets hoger uitkomt.

Tabel 3.13: Overeenkomst tussen oude AVI-toets (versie A) en de nieuwe AVI-toets (versie A en B) bij leerlingen uit de tweede try-out (gepercenteerd naar oud AVI-niveau)

niveau volgens oude toets versie A	niveau nieuwe toets versie A is:				niveau nieuwe toets versie B is:			
	gelijk	lager	hoger	N	gelijk	lager	hoger	N
< 1	89,5		10,5	57	100,0			56
1	56,5	39,1	4,3	23	47,8	47,8	4,3	23
2	40,0	48,0	12,0	25	37,5	35,9	16,7	24
3	30,8	46,2	23,1	13	7,7	76,9	15,4	13
4	34,8	43,5	21,5	23	40,0	44,0	16,0	25
5	25,0	56,3	18,8	16	38,9	38,9	22,2	18
6	20,8	66,7	12,5	24	47,6	33,3	19,0	21
7	37,5	56,3	6,3	32	22,2	69,4	8,3	36
8	42,1	47,4	10,5	38	25,6	55,8	18,7	43
9	20,0	26,4	53,6	110	19,6	23,2	57,1	112
> 9	57,9	42,1		19	81,0	19,0		21
totaal	41,0	35,5	23,4	380	41,3	34,7	24,0	392

Op voorhand hadden we geen grote overeenstemming in resultaten verwacht, maar het gevonden percentage overeenkomstige niveaus is wel erg laag. In de praktijk zal dit gevolgen hebben wanneer scholen overstappen van de oude naar de nieuwe AVI-toets: leerlingen zullen nogal eens terugvallen in niveau of niet vooruitgaan in niveau. Hoe groot dit praktische probleem zal zijn valt niet te voorspellen, mede omdat er ook nog zeer verschillende normeringen voor de oude AVI-toetskaarten in omloop zijn.

3.4 De normering

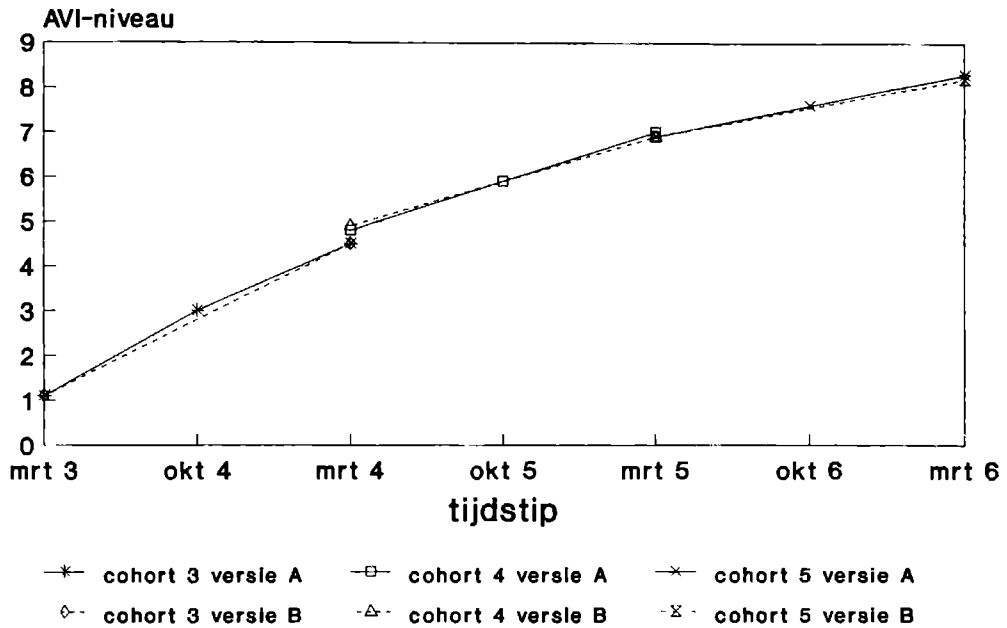
Nu we beschikken over de grenswaarden voor de toetskaarten kan van elke leerling het AVI-niveau worden vastgesteld. Daarmee kunnen we per tijdstip de verdeling van de leerlingen over de AVI-niveaus bepalen (onderzoeksvraag 3). Voor maart van elke groep betreft die zowel de indeling volgens versie A als die volgens versie B, voor oktober alleen volgens versie A. In tabel 3.14 staan de resultaten weergegeven. Hierbij moeten we wel opmerken dat de steekproef op basis waarvan deze gegevens zijn bepaald door uitval niet volledig representatief is voor alle Nederlandse basisscholen. Aarnoutse et al (1996)

hebben wegingsfactoren berekend om voor deze uitval te corrigeren, maar deze waren nog niet beschikbaar ten tijde van het normeringsonderzoek van de AVI-toetskaarten.

Tabel 3.14: Verdeling van de AVI-niveaus per tijdstip per versie (aantal en percentage)

tijdstip	versie	AVI-niveau											
		< 1	1	2	3	4	5	6	7	8	9	> 9	
mrt grp 3	A	N	459	272	149	91	25	30	5	4	2	0	0
		%	44,3	26,2	14,4	8,8	2,4	2,9	0,5	0,4	0,2	0,0	0,0
	B	N	446	256	150	90	35	23	6	4	1	0	0
		%	44,1	25,3	14,8	8,9	3,5	2,3	0,6	0,4	0,1	0,0	0,0
okt grp 4	A	N	87	149	184	177	130	136	45	20	18	2	2
	%	9,2	15,7	19,4	18,6	13,7	14,3	4,7	2,1	1,9	0,2	0,2	
mrt grp 4	A	N	22	57	81	166	138	158	134	113	58	24	17
		%	2,3	5,9	8,4	17,1	14,3	16,3	13,8	11,7	6,0	2,5	1,8
	B	N	22	57	98	146	153	125	158	104	63	27	22
		%	2,3	5,8	10,1	15,0	15,7	12,8	16,2	10,7	6,5	2,8	2,3
okt grp 5	A	N	5	23	36	77	99	106	152	130	106	50	66
	%	0,6	2,7	4,2	9,1	11,6	12,5	17,9	15,3	12,5	5,9	7,8	
mrt grp 5	A	N	2	0	10	35	62	96	149	162	141	98	115
		%	0,2	0,0	1,1	4,0	7,1	11,0	17,1	18,6	16,2	11,3	13,2
	B	N	1	3	13	32	79	83	158	153	139	81	139
		%	0,1	0,3	1,5	3,6	9,0	9,4	17,9	17,4	15,8	9,2	15,8
okt grp 6	A	N	0	1	5	17	37	53	101	150	162	95	184
	%	0,0	0,1	0,6	2,1	4,6	6,6	12,5	18,6	20,1	11,8	22,9	
mrt grp 6	A	N	0	2	1	6	14	38	49	122	155	99	284
		%	0,0	0,3	0,1	0,8	1,8	4,9	6,4	15,8	20,1	12,9	36,9
	B	N	0	0	3	4	24	36	87	106	148	93	294
		%	0,0	0,0	0,4	0,5	3,0	4,5	10,9	13,3	18,6	11,7	37,0

In de tabel is duidelijk de verschuiving in de tijd waarneembaar van de verdeling naar AVI-niveau. In maart van groep 3 beheerst 44% van de leerlingen AVI-niveau 1 nog niet, terwijl in oktober van groep 4 dat nog slechts geldt voor 9% van de leerlingen. Uiteindelijk beheerst in maart van groep 6 zo'n 50% van de leerlingen AVI-niveau 9 (percentages van niveau 9 en >9 opgeteld). De verschuiving van de AVI-niveaus in de tijd lijkt zeer gelijkmatig te verlopen, zonder grote sprongsgewijze verschuivingen. Wel zien we dat de groei van de technische leesvaardigheid, weergegeven in AVI-niveaus, het sterkst is aan het begin van de basisschoolperiode. Dit is duidelijk zichtbaar in figuur 3.9, waarin we het gemiddelde AVI-niveau hebben weergegeven per tijdstip per cohort.

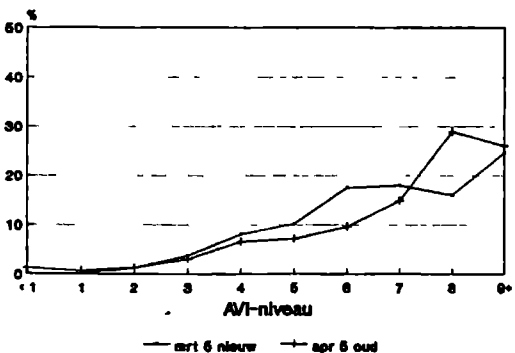
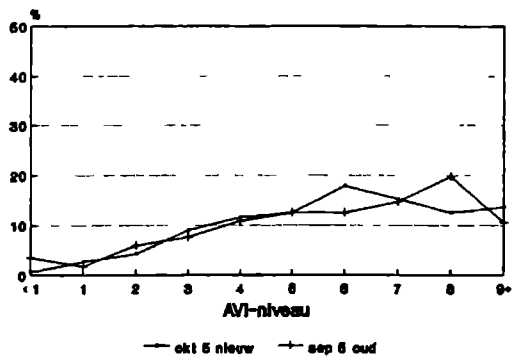
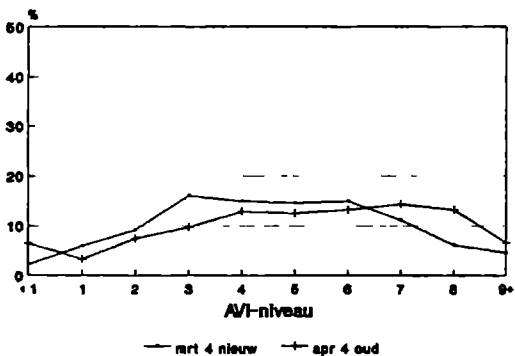
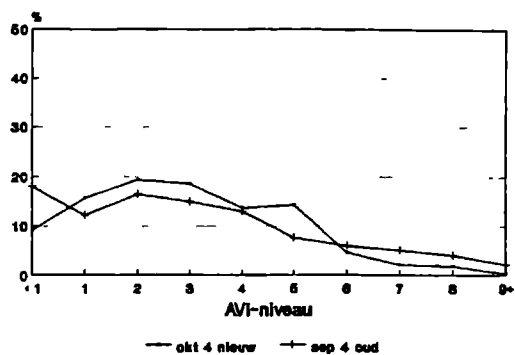
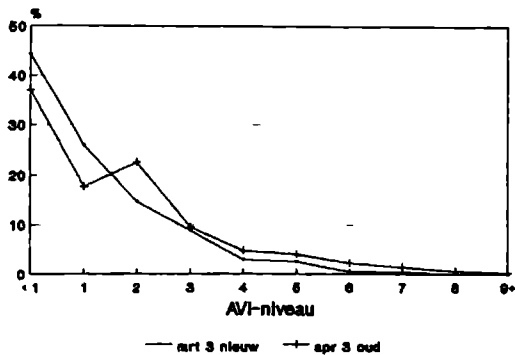


Figuur 3.9: Gemiddeld AVI-niveau per cohort per tijdstip voor versie A en versie B

Uit tabel 3.14 wordt tevens duidelijk dat de verschillen in de technische leesvaardigheid, gemeten met de AVI-toets, van leerlingen binnen een groep groot zijn. Zo beheerst bijvoorbeeld in oktober van groep 5 zo'n 16% AVI-niveau 4 nog niet, terwijl zo'n 14% AVI-niveau 9 (of hoger) beheerst. Ook in andere onderzoeken is zo'n grote spreiding in technische leesvaardigheid gevonden, waarbij ook andere toetsen voor technisch lezen zijn gebruikt (Boland, 1991; Verhoeven & Gillijns, 1994).

Tot slot kan in tabel 3.14 de verdeling over de AVI-niveaus volgens versie A en volgens versie B worden vergeleken in maart van iedere jaargroep. Deze verdelingen komen grotendeels met elkaar overeen.

In figuur 3.10 vergelijken we de verdeling volgens de nieuwe toetskaarten met de verdeling van de oude AVI-toets per tijdstip. Voor de verdeling van de oude AVI-toets maken we gebruik van de gegevens uit Dickhout-Rutten (1983). De meetpunten in de basisschool zijn in dat onderzoek niet geheel gelijk aan die in ons onderzoek: april in plaats van maart en september in plaats van oktober. Aangezien Dickhout-Rutten geen onderscheid heeft gemaakt in niveau 9 en hoger dan niveau 9, hebben we deze twee categorieën in onze gegevens samengevoegd. Tevens hebben we voor de gegevens uit maart een gemiddelde verdeling berekend van de A- en B-versie.



Figuur 3.10: Vergelijking van de percentuele verdeling van de AVI-niveaus volgens de oude (uit Dickhout-Rutten, 1983) en de nieuwe toetskaarten per tijdstip

De verdelingen lopen op de diverse tijdstippen niet gelijk. Deels is dat uiteraard te verklaren door het tijdsverschil van een maand. Zo beheerst in de steekproef van Dickhout-Rutten in april van groep 3 37% AVI-niveau 1 nog niet, tegen 44% in maart van groep 3 in dit onderzoek. Over het algemeen valt uit de grafieken af te lezen dat de oude AVI-niveaus in april van groep 3, 4 en 5 wat hoger liggen dan de nieuwe niveaus in maart van groep 3, 4 en 5. In het voorjaar van groep 5 zijn de resultaten voor niveau 9 en hoger echter wel vergelijkbaar: 26% in april van groep 5 en 25% in maart van groep 5. Wanneer de verschillen in niveaus alleen door het tijdsverschil verklaard zouden worden, dan zouden we in de najaarsmetingen het omgekeerde moeten zien: wat hogere niveaus met de nieuwe AVI-toets. Dat is echter niet over de gehele linie het geval: ook hier zien we dat op een aantal plaatsen het niveau volgens de oude AVI-toets hoger ligt. Dit is, gezien de eerder gevonden resultaten bij de try-out gegevens niet verwonderlijk: de nieuwe AVI-toets is wat strenger, met name tussen de niveaus 5 en 8.

Om scholen de mogelijkheid te bieden snel een indicatie te krijgen van de leesprestaties van de individuele leerlingen hebben we voor elk van de zeven tijdstippen vier prestatiegroepen gecreëerd:

- 1 goed tot zeer goed ($\pm 25\%$ hoogst scorende leerlingen);
- 2 voldoende tot goed ($\pm 25\%$ net boven het landelijk gemiddelde);
- 3 matig tot voldoende ($\pm 25\%$ net onder het landelijk gemiddelde);
- 4 zwak tot matig ($\pm 25\%$ laagst scorende leerlingen).

Bij bestudering van tabel 3.14 blijkt dat deze categorieën niet altijd gelijkmatig gevuld kunnen worden. Bijvoorbeeld in maart van groep 3 beheerst zo'n 44% van de leerlingen AVI-niveau 1 nog niet. Getracht is zo goed mogelijk die 25% in de vier categorieën te benaderen. Tabel 3.15 geeft de uiteindelijke indeling weer (zie ook Visser et al, 1996), waarbij we gebruik hebben gemaakt van de gemiddelde verdeling van de A- en de B-versie.

Tabel 3.15: Indeling in vier prestatiegroepen per tijdstip

Niveau	groep 3		groep 4		groep 5		groep 6	
	maart	oktober	maart	oktober	maart	oktober	maart	oktober
goed tot zeer goed	3 of hoger	4 of hoger	7 of hoger	8 of hoger	9 of hoger	9 of hoger	hoger dan 9	
voldoende tot goed	2	3	5-6	6-7	7-8	8	9	
matig tot voldoende	1	2	3-4	4-5	6	7	8	
zwak tot matig	lager dan 1	1 of lager	2 of lager	3 of lager	5 of lager	6 of lager	7 of lager	

De gegevens uit deze tabel kunnen als volgt worden geïnterpreteerd. Wanneer een leerling in maart van groep 3 wordt ingedeeld in AVI-niveau 2, dan behoort deze leerling tot de categorie 'voldoende tot goed'. Dit betekent dat die leerling tot die 25% van de leerlingen behoort die net boven het landelijk gemiddelde scoren. Een leerling met datzelfde niveau in maart van groep 4 behoort echter tot de categorie 'zwak tot matig', ofwel tot de 25% laagst scorende leerlingen.

Door middel van deze tabellen kan vrij snel worden nagegaan hoe goed de prestaties van een individuele leerling zijn. De resultaten kunnen een eerste indicatie geven voor een vertraagde leesontwikkeling. Leerlingen uit de categorie 'zwak tot matig' gelden als risicoleerlingen die goed in de gaten gehouden moeten worden.

3.5 Samenvatting en conclusies

In dit hoofdstuk hebben we verslag gedaan van de schaling en normering van de AVI-toets. We hebben geprobeerd de toetskaarten te schalen op basis van de leestijden met behulp van het door Jansen aangepaste Rasch-model voor leestijden. Vervolgens zijn de indelingscriteria voor de AVI-niveaus bepaald.

Het gebruikte model lijkt goed bruikbaar voor het schalen van een test als de AVI-toets: een toets bestaande uit meerdere leeskaarten waarvan slechts een deel wordt gelezen door een leerling. Momenteel zijn echter nog geen gegevens beschikbaar over de stabiliteit van de uitkomsten van het model en over de geldigheid van de assumpties.

De toepassing van het model leverde belangrijke resultaten op voor de beantwoording van de eerste onderzoeksvraag: lopen de AVI-toetskaarten op in moeilijkheidsgraad? Het bleek dat de moeilijkheidsgraad van de toetskaarten geschat met het model niet volledig overeenkomt met de ordening op basis van de AVI-niveau criteria (zie bijlage 1). De verschillen in moeilijkheidsgraad tussen de toetskaarten van de verschillende niveaus (vooral de hogere) zijn soms erg klein. Dit bevestigt ons vermoeden (zie paragraaf 2.1.2) dat de verschillen in de leesindices tussen de aangrenzende AVI-niveaus te klein zijn om een goed onderscheid te kunnen maken tussen de niveaus. Ook ander onderzoek naar leesbaarheidsformules geeft aan dat leesbaarheidsindices hooguit indicatoren zijn van de moeilijkheidsgraad van teksten (zie Klare, 1984). In hoofdstuk 5 zullen we nader ingaan op het gevolg van deze resultaten voor de AVI-toetskaarten.

Bij het bepalen van de grenswaarden voor de leestijden per kaart hebben we rekening gehouden met het gegeven dat de volgorde in moeilijkheidsgraad van de AVI-toetskaarten niet geheel conform de verwachting is. Daarbij zijn we namelijk uitgegaan van de grenzen op de zogenaamde theta-schaal voor de leesvaardigheid. De grens-theta's bepalen telkens voor iedere kaart de kritieke leestijden. De leestijden voor een iets makkelijker kaart worden dan simpelweg wat strenger. Zo is bijvoorbeeld kaart 8A moeilijker dan kaart 8B volgens de uitkomsten van het model. De grenswaarden voor de leestijden zijn daarom voor kaart 8B duidelijk strenger dan voor 8A: 95 seconden ten opzichte van 110 seconden voor de voldoende-grens. Een gevolg van deze procedure is wel dat voor de kaarten van niveau 6, 8 en 9 het verschil tussen de leestijden voor de voldoende- en goed-grens slechts 10 seconden bedraagt en voor de niveaus 5 en 7 slechts 15 seconden.

Vrij kleine verschillen in leestijden kunnen daardoor leiden tot een verschil van twee AVI-niveaus. Door deze kleine intervallen wordt ook de kans op inconsistenties groter.

Het uiteindelijke doel van het onderzoek was het bepalen van de grenswaarden die het AVI-niveau aangeven dat een leerling beheerst. We zijn nagegaan in hoeverre de uitkomsten van de verschillende toetskaarten consistent waren (onderzoeksvraag 2).

De eerste controle betrof het vergelijken van de uitkomsten van twee kaarten van hetzelfde niveau: de kans op een verschillend resultaat van twee kaarten van hetzelfde AVI-niveau. Het percentage consistente uitkomsten was het hoogst voor de niveaus 1, 2 en 3 (75%-82%) en voor 4, 8 en 9 (69%-70%). Met name voor de niveaus 5, 6 en 7 was de kans op inconsistenties groot (36%-38%). Het is echter niet zo dat kaarten van versie B consistent lagere uitkomsten geven dan kaarten van versie A, of omgekeerd. Een deel van de inconsistente uitkomsten valt toe te schrijven aan toeval: de prestaties van leerlingen zijn niet altijd constant, maar wisselen van het ene moment op het andere. Daarnaast worden de inconsistenties deels veroorzaakt door het feit dat de grenzen voor de leestijden zijn afgerond tot een veelvoud van 5 seconden.

De tweede controle betrof die op de uitkomsten van twee opeenvolgende toetskaarten. De uitkomsten waren in 68% tot 89% van de gevallen consistent. Twee typen inconsistenties konden worden onderscheiden. Inconsistenties van het type I betroffen de gevallen waarin een kaart van het hogere niveau werd beheerst, terwijl dat hogere niveau volgens de lagere kaart niet werd beheerst. Type II omvatte de gevallen waarin een hogere kaart nog niet beheerst werd, terwijl op de lagere kaart de score goed werd behaald. De kans op een dergelijk resultaat was voor vrijwel alle paren van opeenvolgende kaarten redelijk klein. Gezien de opzet van de afname van de AVI-toetskaarten, is het eerste type ernstiger. Bij de toetsafname wordt namelijk telkens een kaart hoger aangeboden totdat op een kaart de score voldoende wordt behaald. Een type-I inconsistentie betekent dat bij doortoetsen zou blijken dat de leerling toch op een hoger AVI-niveau kan lezen. Deze inconsistentie vonden we vooral bij die paren van opeenvolgende kaarten waarvan de empirisch vastgestelde moeilijkheidsgraden even groot of omgekeerd in volgorde waren. Tot slot is de overeenkomst bepaald tussen het uiteindelijke AVI-niveau volgens versie A en versie B. De overeenkomst is enigszins teleurstellend: 56,1%. In die gevallen waarin de uitkomsten niet overeenstemmen, betreft het meestal een verschil van één AVI-niveau. Na controle voor het hebben gelezen van de juiste toetskaarten bleek de overeenstemming enigszins hoger: 62,2%. De consistentie van de uitkomsten zou verder verhoogd kunnen worden tot 67,9% door de procedure iets te versoepelen, namelijk door uitsluitend de voldoende-grenzen te hanteren. Deze procedure wordt echter niet gehanteerd bij het afnemen van de AVI-toets.

Nadat we de interne consistentie van de nieuwe toets zijn nagegaan, hebben we de overeenstemming tussen de oude en de nieuwe toets bepaald op basis van de gegevens uit de tweede try-out. Deze overeenstemming bleek niet erg groot te zijn. In veel gevallen viel het niveau volgens de nieuwe toets lager uit. Volledige overeenstemming is overigens nooit de bedoeling geweest. Immers, de verdeling van de oude AVI-niveaus liet nogal te wensen over. De matige overeenstemming heeft echter gevolgen voor de

praktijk. Het is namelijk niet ondenkbaar dat leerlingen volgens de nieuwe AVI-toets op een lager niveau ingedeeld zouden moeten worden of dat ze niet vooruit zijn gegaan.

Tot slot hebben we ten behoeve van de beantwoording van de derde onderzoeksvraag de verdeling van de AVI-niveaus bepaald op diverse tijdstippen in de basisschoolperiode. Het aangepaste Rasch-model geeft, naast een schatting van de moeilijkheidsgraad van de toetskaarten, ook een schatting van de leesvaardigheid van de leerlingen, theta genoemd. Deze theta's lijken een goed beeld te kunnen geven van de technische leesvaardigheid van de leerlingen, gemeten uiteraard met de AVI-toets. De gemiddelde theta-waarden geven duidelijk een toename van de leesvaardigheid gedurende de basisschoolperiode te zien. De groei in technische leesvaardigheid is het sterkst aan het begin van het leesonderwijs en neemt af in de hogere jaargroepen.

Dezelfde trend is zichtbaar bij de AVI-niveaus. In de loop van het basisonderwijs neemt het gemiddelde AVI-niveau van de leerlingen toe, waarbij de groei het sterkst is tussen maart groep 3 en maart groep 4.

4 VALIDITEITSONDERZOEK

Het onderzoek naar de validiteit van de AVI-toetskaarten staat centraal in dit hoofdstuk. In dit onderzoek richten we ons op de onderzoeksvragen 4 tot en met 6, zoals geformuleerd in de inleiding van dit proefschrift:

- 4 Kunnen de A- en B-versie van de AVI-toets als paralleltests worden beschouwd?
- 5 Wat is de relatie tussen de AVI-toets en de Eén-Minuut-Test (EMT) van Brus en Voeten (1973)?
- 6 Wat is de relatie tussen de AVI-toets en toetsen voor begrijpend lezen en woordenschat?

De vierde onderzoeksvraag betreft feitelijk nog de betrouwbaarheid van de AVI-toets, maar we presenteren deze vraag in dit hoofdstuk omdat we die met dezelfde onderzoeksmethode beantwoorden als de overige vragen.

Het hoofdstuk is als volgt opgebouwd. In paragraaf 4.1 beschrijven we de onderzoeksofzet waarin we zowel de gebruikte data als de analyseprocedure uiteenzetten. Vervolgens bespreken we in paragraaf 4.2 de relatie tussen de AVI-toetskaarten en de Eén-Minuut-Test (onderzoeksvraag 5), waarin we tevens onderzoeksvraag 4 zullen beantwoorden. De relatie tussen de AVI-toets en de toetsen voor begrijpend lezen en woordenschat (onderzoeksvraag 6) komt aan bod in de paragrafen 4.3 en 4.4. In de paragraaf 4.3 wordt deze relatie per meetperiode onderzocht, in paragraaf 4.4 longitudinaal. Paragraaf 4.5 bevat tot slot de conclusies ten aanzien van de drie onderzoeksvragen.

4.1 De onderzoeksofzet

4.1.1 De steekproef

Voor het validiteitsonderzoek hebben we dezelfde steekproef en dezelfde perioden (voorjaar 1992, najaar 1992 en voorjaar 1993) gebruikt als bij de schaling en normering zoals besproken in hoofdstuk 3. Voor de samenstelling van de bestanden zijn we uitgegaan van de cohortopbouw zoals die binnen het project van de Vakgroep Onderwijskunde Nijmegen wordt gehanteerd (p. 89 in Aarnoutse, Van Leeuwe, Voeten, Van Kan & Oud, 1996). In dat project heeft men ervoor gekozen om maart 1993 als uitgangspunt te nemen voor het vaststellen van het cohort waarin een leerling thuis hoort. Dit heeft men gedaan om het probleem met zittenblijvers op te lossen; deze leerlingen doen immers een jaar over en verschuiven als het ware van cohort. Het probleem is opgelost door de leerlingen tot dat cohort te rekenen waarin ze in maart 1993 zitten. Degenen die vóór maart 1993 zijn blijven zitten, vallen dan niet onder het cohort waarin ze bij de start van het project zaten. Bijvoorbeeld een leerling die in maart 1992 nog in groep 4 zat, en dus ook in cohort 4, maar vervolgens is blijven zitten, is in cohort 3 geplaatst.

Voor het validiteitsonderzoek hebben we de toetsgegevens gebruikt op de terreinen technisch lezen, begrijpend lezen en woordenschat. Voor technisch lezen zijn in elke periode de A- en B-versie van de Eén-Minuut-Test en de A-versie van de AVI-toets afgenomen. In maart 1992 en maart 1993 is tevens de B-versie van de AVI-toets afgenomen. Als scores

op de AVI-toets zijn de AVI-niveaus gebruikt, die zijn bepaald volgens de procedure die beschreven staat in hoofdstuk 3. Voor elke periode zijn twee toetsen beschikbaar voor begrijpend lezen en twee voor woordenschat. Deze toetsen verschillen per tijdstip van meting; de metingen voor twee opeenvolgende perioden hebben steeds één toets gemeenschappelijk. In bijlage 11 staan voor begrijpend lezen en woordenschat de bijbehorende toetsen weergegeven.

Zoals in de meeste onderzoeken is ook hier sprake van ontbrekende gegevens. In dit onderzoek is sprake van drie soorten uitval. Op de eerste plaats hebben we te maken met uitval van scores: sommige leerlingen hebben één of meer toetsen niet gemaakt. Daarnaast is sprake van uitval van leerlingen doordat ze blijven zitten, de school verlaten of nieuw op school komen. Tot slot is sprake van uitval van scholen wanneer een school besluit de deelname aan het onderzoek te stoppen. Bij de analyses per meetperiode hebben we uitsluitend te maken met uitval van scores, bij longitudinale analyses vormt daarnaast ook de uitval van leerlingen en scholen een probleem.

Ten aanzien van ontbrekende gegevens maakt Rubin (1976, zie Allison, 1987; Rovine & Delaney, 1990) onderscheid tussen 'missing at random' en 'observed at random'. Gegevens van een variabele zijn 'missing at random' als het patroon van ontbrekende gegevens niet afhangt van de waarde van die variabele. Zo zijn de gegevens van de AVI-toets niet 'missing at random' als vooral de AVI-niveaus ontbreken van degenen die een laag AVI-niveau hebben. Gegevens van een variabele zijn 'observed at random' als het patroon van ontbrekende gegevens onafhankelijk is van andere geobserveerde variabelen. Zo zijn de gegevens van de AVI-toets niet 'observed at random' als er bijvoorbeeld een verband bestand tussen het ontbreken van een AVI-score en de etnische herkomst van een leerling. Als variabelen zowel 'missing at random' als 'observed at random' zijn, noemt men deze variabelen 'missing completely at random'.

De gegevens in dit onderzoek zijn zeker niet 'observed at random'. Met name voor de gegevens uit het najaar 1992 en het voorjaar 1993 geldt dat het ontbreken van gegevens deels samenhangt met de etnische herkomst en de sociaal-economische status van leerling: scholen in grote steden met veel allochtone leerlingen en Nederlandse achterstandsl leerlingen zijn ondervertegenwoordigd in de steekproef als gevolg van differentiële uitval (Aarnoutse et al, 1996). Aarnoutse et al hebben hiervoor wegingsfactoren berekend, waarvan we in dit onderzoek geen gebruik maken.

Er zijn verschillende methoden beschikbaar om ontbrekende gegevens te schatten waarbij er veelal van uit wordt gegaan dat de gegevens 'missing at random' zijn (Allison, 1987; Rovine & Delaney, 1990). Het is de vraag of de gegevens in het huidige onderzoek 'missing at random' zijn. Zo zijn bijvoorbeeld de zwakkere leerlingen afgevalen doordat ze zijn blijven zitten of zijn verwezen naar het speciaal onderwijs. We zullen verderop nog terugkomen op deze vraag.

Het probleem van de ontbrekende gegevens is feitelijk niet volledig bevredigend op te lossen. Ten behoeve van het validiteitsonderzoek willen we maximaal gebruik maken van de beschikbare gegevens. We hebben ervoor gekozen om de ontbrekende gegevens zoveel mogelijk te schatten. Daarbij hebben we gebruik gemaakt van de imputatie-procedure in PRELIS-2 (Jöreskog & Sörbom, 1993b) waarbij de ontbrekende waarden vervangen wor-

den door reële waarden. Deze reële waarde wordt verkregen uit die gevallen die op de betreffende variabele wel een bekende score hebben en op een set van zogenaamde matchende variabelen precies dezelfde scores hebben als de respondent met de ontbrekende waarde. Voor meer details over deze imputatie-methodiek verwijzen we naar Jöreskog en Sörbom, 1993b).

Om te voorkomen dat bij dat schatten te veel gekapitaliseerd zou worden op kans, hebben we wel enkele beperkingen opgelegd.

Op de eerste plaats hebben we ervoor gekozen de ontbrekende toetsgegevens alleen per meetperiode te schatten. Dat wil zeggen dat we per cohort per periode eventueel ontbrekende gegevens hebben opgevuld, uitsluitend gebruikmakend van de bekende toetsgegevens in de betreffende periode. Niet opgevuld zijn de toetsgegevens die onbekend zijn doordat een leerling in de betreffende periode niet deelnam aan het onderzoek.

Op de tweede plaats hebben we voor een leerling voor een bepaalde periode alleen ontbrekende gegevens geschat als minimaal de helft van de toetsgegevens in die periode voor die leerling wel bekend was.

De derde beperking die we hebben opgelegd is dat een ontbrekende toetsscore alleen geschat mag worden met behulp van daadwerkelijk geobserveerde scores en niet met behulp van andere geschatte scores. Een voorbeeld kan dit duidelijk maken. Stel dat van een leerling de gegevens op toets A en B ontbreken en de scores op de toetsen C tot en met F bekend zijn. In dat geval hebben we eerst de score van toets A geschat gebruikmakend van de toetsscores C tot en met F. Vervolgens hebben we de score voor toets B geschat, wederom alleen met behulp van de scores C tot en met F, en *niet* ook nog met behulp van de geschatte score op toets A. Daarbij hebben we er tevens voor gezorgd dat bij het imputeren alleen gebruik is gemaakt van cases die daadwerkelijk geobserveerde (en dus geen geschatte) scores hadden op de toetsen C tot en met F (en uiteraard op A respectievelijk B).

In tabel 4.1 is per cohort per periode de uiteindelijke bestandsgrootte weergegeven. In het onderzoek voor de onderzoeksvragen 4 en 5 en een deel van 6 (paragrafen 4.2 en 4.3) zijn per cohort per periode de gegevens van die leerlingen gebruikt die na schatting van ontbrekende gegevens op alle toetsen valide waarnemingen hebben.

In het voorjaar van 1992 is de uiteindelijke uitval van leerlingen het grootst: 10%-15%. In het najaar van 1992 en het voorjaar van 1993 is de uitval, na schatting van ontbrekende toetsgegevens, verwaarloosbaar klein.

Tabel 4.1: Uiteindelijke grootte van de opgevulde bestanden per cohort per periode na schatting van ontbrekende gegevens

	oorspronkelijke N	N bij maximaal 4 ontbrekende toetsgegevens	na schatting alle toetsgegevens bekend	
			N	% t.o.v. oorspron- kelijke N
COHORT 3				
voorjaar 1992	1114	1007	954	85,6
najaar 1992	958	953	949	99,1
voorjaar 1993	955	945	944	98,8
COHORT 4				
voorjaar 1992	1026	956	937	91,3
najaar 1992	857	846	846	98,7
voorjaar 1993	856	843	843	98,5
COHORT 5				
voorjaar 1992	956	874	852	89,1
najaar 1992	814	807	806	99,0
voorjaar 1993	812	804	802	98,8

In tabel 4.1 is alleen nog maar de uitval per periode weergegeven. Daarnaast is ook nog sprake van longitudinale uitval, waarmee we te maken hebben bij het onderzoek voor de beantwoording van vraag 6 (paragraaf 4.4). In de longitudinale analyses zijn namelijk alleen de leerlingen betrokken die in alle drie perioden voor alle toetsen valide waarnemingen hebben, na schatting van de ontbrekende gegevens. In hoofdstuk 3 hebben we al aangegeven dat door het afvallen van enkele scholen tussen het voorjaar van 1992 en het najaar van 1992 de longitudinale uitval groot is tussen deze twee perioden. In tabel 4.2 zijn de patronen van bekende gegevens weergegeven, waarin we de geschatte toetsgegevens verdisconteerd hebben.

Tabel 4.2: Aantal leerlingen per patroon van bekende gegevens (na schatting) per cohort

periode			cohort		
voorjaar 1992	najaar 1992	voorjaar 1993	3	4	5
X	X	X	859	796	739
X	X	-	11	8	5
X	-	X	1	1	0
X	-	-	83	132	108
-	X	X	72	38	57
-	X	-	7	4	5
-	-	X	12	8	6
totaal			1.045	987	920

N.B.: X = toetsgegevens beschikbaar

Het eerste patroon betreft de leerlingen van wie we alle benodigde toetsgegevens (eventueel geschat) bezitten over de drie perioden. Dit betreft 859 leerlingen van cohort 3, 796 leerlingen van cohort 4 en 739 leerlingen van cohort 5. De gegevens van deze leerlingen gebruiken we in paragraaf 4.4 bij de ontwikkeling van de longitudinale modellen. Wanneer we deze aantallen relateren aan de oorspronkelijke aantallen in het voorjaar van 1992 (zie de tweede kolom van tabel 4.1) dan is van deze leerlingen nog 77% over (cohort 3: 77,1%; cohort 4: 77,6%; cohort 5: 77,3%). De longitudinale uitval bedraagt tussen het voorjaar van 1992 en het voorjaar van 1993 zo'n 23%, waarvan het merendeel toe te schrijven is aan de uitval van enkele scholen.

We zijn vervolgens nagegaan in hoeverre de leerlingen die in het validiteitsonderzoek zijn betrokken verschillen van de uitvallers, waardoor we deels een antwoord krijgen op de vraag of de gegevens 'missing at random' zijn. De tabellen 4.3 tot en met 4.5 bevatten de descriptieve statistieken van de toetsen voor technisch lezen, begrijpend lezen en woordenschat per cohort per periode, inclusief de geschatte toetsscores. Voor verdeling van een aantal variabelen geldt dat, gezien de scheefheids- en kurtosismaten, deze in grote mate afwijkt van de normaalverdeling. Dit zijn de AVI-toetsen en Lees en Begrijp 1 in cohort 3, voorjaar groep 3 en de Woordenschat van Stijnen in het voorjaar van groep 5 in zowel cohort 4 als cohort 5. Tevens geldt voor nogal wat variabelen dat sprake is van een bodem- of plafondeffect. Een speciaal probleem in verband met de verdelingsassumpties is het kleine aantal waarden op de AVI-toetskaarten. In bijlage 12 staan de descriptieve statistieken weergegeven voor dezelfde toetsen in het oorspronkelijke bestand.

Tabel 4.3: Descriptieve statistieken van de toetsen voor technisch lezen per cohort per periode, opgevuld bestand

	Gemiddelde	Standaardafwijking	Kurtosis	p	Skewness	p	min	max	N
COHORT 3 voorjaar groep 3									
EMT versie A	26,3	12,0	1,09	0,00	0,93	0,00	3	74	954
EMT versie B	24,8	12,0	1,16	0,00	1,10	0,00	3	74	954
AVI versie A	1,2	1,4	2,31	0,00	1,47	0,00	0	8	954
AVI versie B	1,2	1,4	1,84	0,00	1,37	0,00	0	8	954
najaar groep 4									
EMT versie A	31,6	13,5	-0,15	0,16	0,52	0,16	7	81	949
EMT versie B	30,5	13,9	-0,11	0,25	0,62	0,25	6	81	949
AVI versie A	3,0	2,0	-0,07	0,34	0,51	0,34	0	10	949
voorjaar groep 4									
EMT versie A	43,0	13,9	-0,46	0,00	0,17	0,02	11	83	944
EMT versie B	42,0	14,3	-0,59	0,00	0,15	0,03	11	85	944
AVI versie A	4,5	2,1	-0,28	0,02	0,14	0,04	0	10	944
AVI versie B	4,5	2,1	-0,37	0,00	0,17	0,02	0	10	944
COHORT 4 voorjaar groep 4									
EMT versie A	45,0	14,5	-0,46	0,00	0,23	0,01	10	91	937
EMT versie B	43,4	14,8	-0,58	0,00	0,17	0,02	10	89	937
AVI versie A	4,9	2,3	-0,65	0,00	0,05	0,21	0	10	937
AVI versie B	5,0	2,4	-0,69	0,00	0,07	0,14	0	10	937
najaar groep 5									
EMT versie A	52,4	14,8	-0,18	0,13	0,15	0,03	13	103	846
EMT versie B	51,0	14,8	-0,39	0,00	-0,03	0,33	8	100	846
AVI versie A	5,9	2,3	-0,56	0,00	-0,15	0,04	0	10	846
voorjaar groep 5									
EMT A	59,2	13,8	-0,15	0,18	-0,03	0,34	17	102	843
EMT B	58,5	14,2	-0,04	0,44	-0,05	0,23	17	105	843
AVI versie A	7,0	2,1	-0,62	0,00	-0,30	0,00	0	10	843
AVI versie B	6,9	2,2	-0,66	0,00	-0,29	0,00	0	10	843
COHORT 5 voorjaar groep 5									
EMT A	58,9	13,5	0,15	0,18	0,31	0,00	22	110	852
EMT B	57,9	13,4	0,25	0,08	0,23	0,01	21	109	852
AVI versie A	6,9	1,9	-0,52	0,00	-0,21	0,01	2	10	852
AVI versie B	6,9	2,0	-0,56	0,00	-0,19	0,02	2	10	852
najaar groep 6									
EMT versie A	63,9	13,6	-0,12	0,26	0,12	0,06	28	110	806
EMT versie B	62,8	13,7	0,18	0,15	0,18	0,02	24	108	806
AVI versie A	7,6	1,9	-0,26	0,05	-0,55	0,00	1	10	806
voorjaar groep 6									
EMT versie A	69,1	13,4	0,11	0,23	0,17	0,03	32	112	802
EMT versie B	68,3	13,7	0,25	0,08	0,23	0,01	26	114	802
AVI versie A	8,3	1,7	0,59	0,00	-0,94	0,00	1	10	802
AVI versie B	8,2	1,8	-0,20	0,11	-0,75	0,00	2	10	802

Tabel 4.4: Descriptieve statistieken van de toetsen voor begrijpend lezen per cohort per periode, opgevuld bestand

	Gemiddelde	Standaardafwijking	Kurtosis	p	Skewness	p	min	max	N
COHORT 3. voorjaar groep 3									
Lees en Begrijp 1 (Cito)	22,8	3,9	3,57	0,00	-1,66	0,00	3	27	954
Lees en Begrijp 2 (Cito)	21,5	7,4	-0,83	0,00	-0,64	0,00	2	30	954
najaar groep 4									
Lees en Begrijp 2 (Cito)	22,4	6,8	-0,28	0,02	-0,87	0,00	1	30	949
Begrijpend Leestest 4 (Aarmoutse)	21,3	7,5	-0,90	0,00	-0,20	0,01	4	36	949
voorjaar groep 4									
Begrijpend Leestest 4 (Aarmoutse)	26,8	6,2	0,44	0,01	-0,93	0,00	5	36	944
Begrijpend Lezen M3 (Cito)	16,5	5,4	-0,95	0,00	-0,29	0,00	3	25	944
COHORT 4: voorjaar groep 4									
Begrijpend Leestest 4 (Aarmoutse)	27,0	6,1	0,37	0,02	-0,87	0,00	5	36	937
Begrijpend Lezen M3 (Cito)	16,9	5,1	-0,60	0,00	-0,47	0,00	2	25	937
najaar groep 5									
Begrijpend Lezen M3 (Cito)	18,4	5,4	-0,29	0,03	-0,77	0,00	2	25	846
Begrijpend Leestest 5 (Aarmoutse)	18,7	6,6	-0,90	0,00	-0,28	0,01	2	30	846
voorjaar groep 5									
Begrijpend Leestest 5 (Aarmoutse)	22,1	5,7	-0,13	0,23	-0,82	0,00	5	30	843
Begrijpend Lezen M4 (Cito)	17,3	5,2	-0,72	0,00	-0,53	0,00	3	25	843
COHORT 5 voorjaar groep 5									
Begrijpend Leestest 5 (Aarmoutse)	22,6	5,2	0,72	0,00	-1,01	0,00	4	30	852
Begrijpend Lezen M4 (Cito)	17,7	4,6	-0,19	0,12	-0,62	0,00	3	25	852
najaar groep 6									
Begrijpend Lezen M4 (Cito)	19,3	4,6	0,33	0,04	-1,00	0,00	4	25	806
Begrijpend Leestest 6 (Aarmoutse)	20,0	5,4	-0,68	0,00	-0,34	0,00	5	30	806
voorjaar groep 6									
Begrijpend Leestest 6 (Aarmoutse)	22,9	5,2	0,41	0,02	-0,93	0,00	5	30	802
Begrijpend Lezen E4 (Cito)	17,5	4,3	-0,15	0,19	-0,60	0,00	2	25	802

Tabel 4.5: Descriptieve statistieken van de toetsen voor woordenschat per cohort per periode, opgevuld bestand

	Gemiddelde	Standaardafwijking	Kurtosis	p	Skewness	p	min	max	N
COHORT 3: voorjaar groep 3									
Begnppentest Woorden (Aarmoutse)	28,9	6,1	0,41	0,01	-0,81	0,00	9	40	954
Woordenschattest 3 (Aarmoutse)	19,4	4,8	0,30	0,04	-0,94	0,00	3	26	954
najaar groep 4									
Woordenschattest 3 (Aarmoutse)	21,4	4,3	1,89	0,00	-1,44	0,00	3	26	949
Woordenschattest 4 (Aarmoutse)	21,6	7,3	-0,94	0,00	-0,23	0,01	5	35	949
voorjaar groep 4									
Woordenschattest 4 (Aarmoutse)	27,1	6,0	0,07	0,30	-0,87	0,00	8	35	944
Woordenschattest 5 (Aarmoutse)	21,0	6,8	-0,77	0,00	0,05	0,21	5	36	944
COHORT 4: voorjaar groep 4									
Woordenschattest 4 (Aarmoutse)	27,3	5,9	0,72	0,00	-1,02	0,00	5	35	937
Woordenschattest 5 (Aarmoutse)	20,6	6,9	-0,80	0,00	0,04	0,26	4	36	937
najaar groep 5									
Woordenschattest 5 (Aarmoutse)	23,8	6,9	-0,78	0,00	-0,31	0,00	6	36	846
Woordenschattest 3 (Stijnen)	38,8	8,6	0,04	0,38	-0,65	0,00	6	55	846
voorjaar groep 5									
Woordenschattest 3 (Stijnen)	43,7	7,6	3,04	0,00	-1,45	0,00	4	55	843
Woordenschattest 6 (Aarmoutse)	20,1	5,1	-0,32	0,02	-0,15	0,03	5	33	843
COHORT 5: voorjaar groep 5									
Woordenschattest 3 (Stijnen)	43,3	8,1	4,72	0,00	-1,82	0,00	5	55	852
Woordenschattest 6 (Aarmoutse)	19,5	5,2	-0,40	0,00	0,02	0,35	5	32	852
najaar groep 6									
Woordenschattest 6 (Aarmoutse)	21,4	5,6	-0,64	0,00	-0,24	0,01	6	33	806
Woordenschattest 4 (Stijnen)	40,1	9,1	0,15	0,20	-0,61	0,00	5	68	806
voorjaar groep 6									
Woordenschattest 4 (Stijnen)	44,6	8,1	1,81	0,00	-1,05	0,00	8	60	802
Woordenschattest 5 (Stijnen)	38,9	8,5	0,01	0,44	-0,41	0,00	12	59	802

Als eerste hebben we met behulp van t-toetsen getoetst of de gemiddelden in *het opgevolde bestand* afwijken van de gemiddelden in het *oorspronkelijke bestand*. Daarbij hebben we deze gemiddelden opgevat als zijnde afkomstig van twee verschillende groepen. Alleen in cohort 3 bleek sprake te zijn van verschillen in gemiddelden tussen het oorspronkelijke en het opgevolde bestand, en wel voor de Eén-Minuut-Test (EMT) versie A en B in het voorjaar van 1992 (de eerste periode). Voor beide toetsen geldt dat het gemiddelde in het opgevolde bestand hoger is dan in het oorspronkelijke bestand. Uit deze resultaten mogen we afleiden dat het opvullen van de ontbrekende toetsgegevens niet tot een verandering van de gemiddelde toetsscores heeft geleid.

Vervolgens zijn we nagegaan of er sprake is van verschillen tussen leerlingen die *wel* en leerlingen die *niet in de analyses per periode zijn betrokken*, in het vervolg de blijvers en de uitvallers genaamd. Daarvoor hebben we wederom per periode per cohort t-toetsen uitgevoerd en wel op de oorspronkelijke gegevens van de toetsen voor technisch lezen. We beschikten voor de uitvallers niet over de toetsgegevens van begrijpend lezen en woordenschat. In tabel 4.6 hebben we de gegevens van beide groepen naast elkaar gezet. Alleen de gegevens van het voorjaar 1992 zijn in de tabel opgenomen, omdat voor de andere perioden het aantal uitvallers gering was. In de meeste gevallen verschilden de varianties tussen beide groepen, waardoor in de tabel de t-toets voor ongelijke varianties is weergegeven.

Tabel 4.6: Toetsing van de verschillen tussen blijvers en uitvallers voor de analyses per periode

	blijvers			uitvallers			t-toets		
	Gemid- delde	Standaard- afwijking	N	Gemid- delde	Standaard- afwijking	N	t	df	p
COHORT 3: voorjaar groep 3									
EMT versie A	25,9	11,8	802	15,9	11,4	69	6,97	81,2	0,00
EMT versie B	24,3	11,8	804	13,7	5,6	70	13,59	132,5	0,00
AVI versie A	1,2	1,4	904	0,3	0,7	134	11,63	323,0	0,00
AVI versie B	1,3	1,4	884	0,4	0,7	128	11,20	306,3	0,00
COHORT 4: voorjaar groep 4									
EMT versie A	44,8	14,5	898	44,3	13,1	26	0,16	922	0,87
EMT versie B	43,2	14,7	900	41,8	13,2	26	0,49	924	0,62
AVI versie A	4,9	2,3	911	4,6	1,9	67	1,04	80,5	0,30
AVI versie B	5,0	2,4	907	4,8	2,1	66	0,69	971	0,49
COHORT 5: voorjaar groep 5									
EMT versie A	58,9	13,6	832	57,3	14,3	16	0,47	846	0,64
EMT versie B	57,9	13,4	820	56,4	16,1	15	0,42	833	0,67
AVI versie A	6,9	1,9	852	7,4	2,8	65	-2,10	915	0,04
AVI versie B	6,9	2,0	853	7,7	1,8	64	-3,09	915	0,00

Uit tabel 4.6 kunnen we aflezen dat in *cohort 3* de blijvers en de uitvallers verschillen op alle toetsen voor technisch lezen in het voorjaar van groep 3 (periode 1). De blijvers scoorden duidelijk beter dan de uitvallers: een verschil van circa 10 woorden op de EMT en één niveau op de AVI-toets. Blijvers en uitvallers van *cohort 4* verschilden niet significant van elkaar op de toetsen voor technisch lezen. In *cohort 5* vonden we verschillen op de beide versies van de AVI-toets in het voorjaar van groep 5, waarbij de gemiddelden van de uitvallers juist hoger zijn.

Tot slot zijn we in het opgevulde bestand per cohort nagegaan in hoeverre sprake is van verschillen in toetsscores tussen degenen van wie *in alle drie perioden de toetsgegevens beschikbaar* zijn (en die dus bij de longitudinale analyses betrokken worden) en degenen van wie de toetsgegevens van slechts één of twee perioden beschikbaar zijn. In tabel 4.7 zijn de resultaten van de t-toetsen weergegeven.

Tabel 4.7: Toetsing van de verschillen tussen degenen die in de longitudinale analyses worden betrokken en degenen die alleen in de analyses per periode worden betrokken

	longitudinale steekproef			uitvallers			t-toets		
	Gemid- delde	Standaard- afwijking	N	Gemid- delde	Standaard- afwijking	N	t	df	p
COHORT 3. voorjaar groep 3									
EMT versie A	26,2	11,9	859	26,4	13,0	95	-0,14	952	0,89
EMT versie B	24,7	11,9	859	25,5	13,2	95	-0,57	952	0,57
AVI versie A	1,2	1,4	859	1,4	1,4	95	-1,27	952	0,21
AVI versie B	1,2	1,4	859	1,6	1,4	95	-2,50	952	0,01
Lees en Begrijp 1	22,7	4,0	859	23,3	3,6	95	-1,33	952	0,19
Lees en Begrijp 2	21,3	7,4	859	22,8	6,7	95	-1,88	952	0,06
Begrijppentest woorden	28,8	6,1	859	29,6	5,7	95	-1,22	952	0,22
Woordenschat 3 (Aam)	19,3	4,8	859	20,2	4,5	95	-1,83	952	0,07
najaar groep 4									
EMT versie A	31,7	13,7	859	31,3	11,9	90	0,27	115,2	0,79
EMT versie B	30,5	14,1	859	30,1	12,2	90	0,29	115,3	0,77
AVI versie A	3,0	2,0	859	3,2	1,9	90	-1,04	947	0,30
Lees en Begrijp 2	22,4	6,9	859	21,7	6,2	90	0,97	947	0,33
Begrijpend Leestest 4	21,2	7,5	859	21,9	7,2	90	-0,83	947	0,41
Woordenschat 3 (Aam)	21,4	4,3	859	21,7	3,5	90	-0,88	119,1	0,38
Woordenschat 4 (Aam)	21,5	7,4	859	22,8	6,1	90	-1,90	117,9	0,06

	longitudinale steekproef			uitvallers			t-toets		
	Gemid- delde	Standaard- afwijking	N	Gemid- delde	Standaard- afwijking	N	t	df	p
voorjaar groep 4									
EMT versie A	43,1	14,0	859	42,2	12,8	85	0,60	942	0,55
EMT versie B	42,1	14,4	859	41,4	13,9	85	0,37	942	0,71
AVI versie A	4,5	2,1	859	4,8	2,2	85	-1,08	942	0,28
AVI versie B	4,5	2,1	859	4,3	2,2	85	1,02	942	0,31
Begrijpend Leestest 4	26,7	6,3	859	27,5	5,5	85	-1,03	942	0,31
Begrijpend Lezen M3	16,5	5,4	859	16,4	4,9	85	0,07	942	0,95
Woordenschat 4 (Aarn)	27,1	6,1	859	27,9	5,3	85	-1,15	942	0,25
Woordenschat 5 (Aarn)	21,0	6,8	859	21,2	6,5	85	-0,37	942	0,71
COHORT 4 voorjaar groep 4									
EMT versie A	45,5	14,3	796	41,8	15,6	141	2,80	935	0,01
EMT versie B	44,1	14,5	796	39,8	16,0	141	2,96	182,5	0,00
AVI versie A	5,0	2,2	796	4,2	2,5	141	3,67	180,9	0,00
AVI versie B	5,1	2,3	796	4,4	2,6	141	3,21	179,9	0,00
Begrijpend Leestest 4	27,2	6,0	796	25,8	6,6	141	2,62	935	0,01
Begrijpend Lezen M3	17,0	5,1	796	16,2	5,1	141	1,77	935	0,08
Woordenschat 4 (Aarn)	27,5	5,8	796	26,2	6,5	141	2,45	935	0,02
Woordenschat 5 (Aarn)	20,8	7,0	796	19,3	6,6	141	2,32	935	0,02
najaar groep 5									
EMT versie A	52,4	14,6	796	51,3	17,4	50	0,46	53,4	0,65
EMT versie B	51,1	14,7	796	50,0	16,3	50	0,71	844	0,48
AVI versie A	6,0	2,3	796	5,6	2,5	50	1,13	844	0,26
Begrijpend Lezen M3	18,4	5,4	796	17,3	5,1	50	1,42	844	0,16
Begrijpend Leestest 5	18,7	6,6	796	17,6	6,6	50	1,19	844	0,24
Woordenschat 5 (Aarn)	23,8	6,9	796	23,4	6,2	50	0,44	844	0,66
Woordenschat 3 (Stijn)	38,9	8,6	796	37,8	9,3	50	0,87	844	0,39
voorjaar groep 5									
EMT versie A	59,3	13,7	796	56,6	15,7	47	1,32	841	0,19
EMT versie B	58,6	14,1	796	56,4	15,3	47	1,02	841	0,31
AVI versie A	7,1	2,1	796	6,7	2,3	47	1,06	841	0,29
AVI versie B	7,0	2,2	796	6,4	2,4	47	1,79	841	0,08
Begrijpend Leestest 5	22,2	5,6	796	20,6	6,7	47	1,58	49,9	0,12
Begrijpend Lezen M4	17,3	5,1	796	16,8	6,2	47	0,53	49,8	0,60
Woordenschat 3 (Stijn)	43,7	7,5	796	42,5	8,7	47	1,05	841	0,30
Woordenschat 6 (Aarn)	20,1	5,1	796	19,5	5,4	47	0,71	841	0,48

	longitudinale steekproef			uitvallers			t-toets		
	Gemid- delde	Standaard- afwijking	N	Gemid- delde	Standaard- afwijking	N	t	df	p
COHORT 5. voorjaar groep 5									
EMT versie A	59,0	13,4	739	58,2	14,3	113	0,61	850	0,54
EMT versie B	58,1	13,2	739	56,9	14,1	113	0,86	850	0,39
AVI versie A	7,0	1,9	739	6,6	1,8	113	2,06	850	0,04
AVI versie B	7,0	2,0	739	6,8	2,0	113	1,06	850	0,29
Begrijpend Leestest 5	22,6	5,2	739	22,6	4,7	113	0,14	850	0,89
Begrijpend Lezen M4	17,7	4,6	739	17,8	4,3	113	-0,27	850	0,79
Woordenschat 3(Stijn)	43,3	8,1	739	43,3	8,4	113	0,01	850	0,99
Woordenschat 6 (Aarn)	19,4	5,2	739	20,0	5,2	113	-1,15	850	0,25
najaar groep 6									
EMT versie A	63,8	13,4	739	64,7	15,9	67	-0,54	804	0,59
EMT versie B	62,9	13,4	739	61,8	16,0	67	0,61	804	0,55
AVI versie A	7,6	1,9	739	7,4	2,1	67	0,91	804	0,37
Begrijpend Lezen M4	19,3	4,7	739	19,0	4,2	67	0,58	804	0,57
Begrijpend Leestest 6	19,9	5,4	739	20,8	5,0	67	-1,26	804	0,21
Woordenschat 6 (Aarn)	21,3	5,6	739	22,1	5,6	67	-1,06	804	0,29
Woordenschat 4(Stijn)	40,1	9,0	739	39,8	9,9	67	0,24	804	0,81
voorjaar groep 6									
EMT versie A	69,3	13,3	739	67,0	15,1	63	1,31	800	0,19
EMT versie B	68,6	13,5	739	65,2	15,7	63	1,86	800	0,06
AVI versie A	8,4	1,7	739	7,6	2,1	63	2,84	68,7	0,01
AVI versie B	8,2	1,8	739	7,7	2,1	63	2,37	800	0,02
Begrijpend Leestest 6	22,8	5,2	739	23,8	4,9	63	-1,40	800	0,16
Begrijpend Lezen E3	17,5	4,3	739	17,7	4,2	63	-0,39	800	0,69
Woordenschat 4(Stijn)	44,6	8,0	739	44,5	8,5	63	0,04	800	0,97
Woordenschat 5(Stijn)	38,9	8,5	739	39,5	8,6	63	-0,53	800	0,60

In *cohort 3* verschilt alleen de gemiddelde score op versie B van de AVI-toets in de eerste periode: degenen die in de longitudinale analyses betrokken worden, scoorden gemiddeld wat lager dan de anderen.

In *cohort 4* scoorden degenen die in het longitudinale bestand zijn opgenomen op alle toetsen, met uitzondering van een toets voor begrijpend lezen in de eerste periode beter dan de overige leerlingen. Voor de andere perioden hebben we geen verschillen gevonden in *cohort 4*.

Voor *cohort 5* vonden we dat degenen uit het longitudinale bestand iets beter scoorden op AVI versie A in de eerste periode en op versie A en B in de derde periode.

Uit het bovenstaande kunnen we concluderen dat de gegevens in dit onderzoek niet 'missing at random' zijn, terwijl de procedure waarmee we de ontbrekende gegevens hebben geschat daar wel vanuit gaat. Voor cohort 3 geldt dat leerlingen die in het onderzoek zijn betrokken beter presteren op toetsen voor technisch lezen dan de uitvallers. In cohort 4 vonden we nauwelijks verschillen en in cohort 5 vonden we in de eerste periode dat de uitvallers beter scoorden op beide versies van de AVI-toets. We mogen echter verwachten dat ook de leerlingen uit de cohorten 4 en 5 gemiddeld wat beter presteren dan de totale populatie. We kunnen immers verwachten dat in deze cohorten de selectie van leerlingen al plaats heeft gehad voordat het onderzoek bij hen startte, namelijk tussen groep 3 en groep 4. Voor alle cohorten geldt tevens dat de gegevens niet 'observed at random' zijn aangezien bepaalde typen scholen ondervertegenwoordigd zijn in het onderzoek.

Wat de consequenties van het bovenstaande zijn voor de uitkomsten van het onderzoek is niet duidelijk. Bij de interpretaties zullen we ons in ieder geval moeten realiseren dat de resultaten niet geldig hoeven te zijn voor de gehele populatie.

4.1.2 Analyseprocedure

In het validiteitsonderzoek maken we gebruik van Lisrel-modellen (Lisrel 8; Jöreskog & Sörbom, 1993a). De strategie die we daarbij volgen komt stapsgewijs neer op het volgende:

- 1 specificatie basismodel;
- 2 modelontwikkeling;
- 3 hypothese-toetsing;
- 4 parameterschatting en interpretatie.

Specificatie basismodel

Op basis van theoretische overwegingen hebben we een model gespecificeerd dat als basismodel fungeert voor de analyses (stap 1). Voor iedere onderzoeksvraag is opnieuw een basismodel opgesteld, uiteraard rekening houdend met de resultaten bij de voorgaande onderzoeksvraag. Voor de onderzoeksvragen 4 en 5 betreft het basismodel een cross-sectioneel model met alleen de toetsen voor technisch lezen (zie paragraaf 4.2). Voor onderzoeksvraag 6 is het uiteindelijke model uit paragraaf 4.2 als uitgangspunt genomen voor een nieuw cross-sectioneel basismodel waarin ook de toetsen voor begrijpend lezen en woordenschat zijn opgenomen. Dit model is vervolgens uitgebreid naar een longitudinaal model (zie paragraaf 4.4).

Modelontwikkeling

Als eerste wordt het basismodel ingevoerd in de analyses, gebruikmakend van covariantie-matrices (zie bijlage 19) en indien van toepassing de gemiddelden. In alle Lisrel-analyses is de 'maximum-likelihood'-methode toegepast. Nagegaan wordt hoe goed het model bij de data past. Lisrel geeft daarvoor diverse fit-maten.

De meest gebruikte maat voor modelfit is de χ^2 -test, die de nulhypothese toetst dat de covariantie-matrix van de waargenomen variabelen in de populatie gelijk is aan de op basis van het model geschatte covariantie-matrix. Anders gezegd: de nulhypothese is dat

het gespecificeerde model leidt tot een exacte reproductie van de covariantie-matrix in de populatie.

We zullen ons echter niet beperken tot deze toets, aangezien daaraan nogal wat bezwaren kleven (Bollen & Long, 1993; Hu & Bentler, 1995). Zo is de nul-hypothese in de meeste gevallen zeer streng: in de sociale wetenschappen zal de werkelijkheid hooguit benaderd, maar nooit exact gereproduceerd kunnen worden. Op de tweede plaats wordt bij de test geen rekening gehouden met het onderscheidingsvermogen van de significantietoets, waardoor bij grote steekproeven kleine onbetekenende verschillen al tot een significante χ^2 kunnen leiden. Goede modellen worden dan ten onrechte verworpen. En tegelijkertijd kunnen bij kleine steekproeven modellen ten onrechte geaccepteerd worden door het gebrek aan statistisch onderscheidingsvermogen. Tot slot kan het feit dat de variabelen niet voldoen aan de verdelingsassumpties eveneens leiden tot ten onrechte verwerpen of accepteren van modellen.

Voor de keuze van de maten voor modelfit hebben we ons gebaseerd op Browne en Cudeck (1993), Hu en Bentler (1995), Raftery (1993) en Tanaka (1993). Zij adviseren het gebruik van de 'Bentler scaled chi-square' in plaats van de χ^2 -test. We hebben echter voor de laatste gekozen omdat de eerste niet beschikbaar is in Lisrel 8.

Naast de χ^2 -test zullen we de *Goodness-of-Fit index (GFI)* van elk model presenteren (Jöreskog & Sörbom, 1981 zie Hu & Bentler, 1995). Deze index geeft aan in welke mate de geobserveerde varianties en covarianties worden verklaard door het model, oftewel hoe goed een model de data reproduceert. De maat is analoog aan de R^2 -waarde bij multiple regressie. De waarden van de GFI kunnen lopen van 0 tot 1. Als criterium voor modelfit wordt vaak gekozen voor $GFI > 0,90$.

De GFI is een zogenaamde absolute fit-index. Daarnaast zijn ook diverse incrementele fit-indices ontwikkeld (voor een overzicht zie Hu & Bentler, 1995; Tanaka, 1993). Deze indices geven de proportionele verbetering van de fit weer wanneer het model vergeleken wordt met een baseline-model (veelal het model waarin alle geobserveerde variabelen ongecorrleerd zijn). In dit proefschrift zullen we van ieder model de *Comparative Fit Index (CFI)* weergeven (Bentler, 1990, zie Hu & Bentler, 1995). Deze index kan een waarde aannemen tussen 0 en 1. Hoe groter de waarde van de index, hoe beter de verbetering van de fit door het model vergeleken met het baseline-model. Overigens blijken ook de fit-indices GFI en CFI (en alle andere vergelijkbare indices) in meer of mindere mate te worden beïnvloed door de steekproefgrootte en de complexiteit van het model (Hu & Bentler, 1995).

Een probleem met de verschillende fit-indices is dat complexere modellen (modellen met minder vrijheidsgraden) vaak bevoordeeld worden, terwijl vanuit het oogpunt van bruikbaarheid de modellen juist eenvoudig (spaarzaam) en begrijpelijk zouden moeten zijn. Om een middenweg te vinden tussen eenvoud en goede modelfit presenteerden Browne en Cudeck (1993) enkele maten die rekening houden met het aantal parameters in het model. Een daarvan is de *RMSEA (Root Mean Square Error of Approximation)*. Deze maat geeft het gebrek aan modelfit aan ten opzichte van de covariantie-matrix in de populatie, waarbij de meer eenvoudige modellen beloond worden. De RMSEA is alleen gelijk aan 0 als het model precies past. Browne en Cudeck stellen voor om niet deze exacte fit te toetsen, maar de zogenaamde 'close fit': in hoeverre de RMSEA afwijkt van 0,05. We heb-

ben hierboven al aangegeven dat het toetsen van de exacte fit bij de χ^2 te streng is. Het basismodel zal vervolgens beperkt dan wel uitgebreid worden conform de te toetsen hypothesen.

Hypothese-toetsing

Bij de presentatie van ieder basismodel zal tevens aangegeven worden op welke manier het model gemodificeerd zal worden om bepaalde hypothesen te kunnen toetsen. We presenteren een aantal concurrerende modellen, die meestal genest zijn. Met behulp van de hierboven genoemde fit-indices kunnen we van ieder van de concurrerende modellen bepalen of het model bij de data past. De volgende vraag is dan: welk model verdient de voorkeur?

De meest gebruikte manier om te toetsen welk van twee geneste modellen de voorkeur verdient is via de toetsing van het *verschil in χ^2* tussen deze modellen. Dit verschil is wederom asymptotisch χ^2 -verdeeld, met als aantal vrijheidsgraden het verschil in vrijheidsgraden tussen beide modellen. Een nadeel van deze methode is echter dat complexere modellen bevoordeeld worden (Hoyle & Panter, 1995; Raftery, 1993). Daarom zullen we naast deze toetsing de modellen nog op een andere manier met elkaar vergelijken.

Raftery (1993) geeft een alternatieve manier om modellen te vergelijken en tot een keuze te komen, namelijk via de *Bayesiaanse benadering*. De vraag daarbij is: welke model voorspelt de data beter. Of anders gezegd: bij welk model is het het meest waarschijnlijk dat deze data geobserveerd zijn. In deze benadering worden tevens de meer simpele modellen bevoordeeld. We zullen, conform de benadering van Raftery, bij ieder model het 'Bayesian Information Criterion' BIC_{ks} weergeven, dat model M_k vergelijkt met het verzadigde model (zie formule 4.1). Als $BIC_{ks} < 0$ dan heeft model M_k de voorkeur boven het verzadigde model.

$$BIC_{ks} = \chi_k^2 - df_k \ln[N \cdot p] \quad (4.1)$$

waarbij:

df_k = aantal vrijheidsgraden in model M_k

N = aantal observaties

p = aantal geobserveerde variabelen.

Voor het vergelijken van twee geneste modellen gaf Jeffreys (1961, zie Raftery, 1993) de volgende interpretatie (de modellen M_0 en M_1 , waarbij M_0 het meest eenvoudige model is (1 parameter minder)) van het verschil in BIC_{ks} van beide modellen:

- 1 als $BIC_{0s} < BIC_{1s}$, dan is er geen aanwijzing voor het additionele effect in model M_1 ;
- 2 als $0 \leq (BIC_{0s} - BIC_{1s}) < 4.6$, dan is er een *zwakke aanwijzing* voor M_1 ;
- 3 als $4.6 \leq (BIC_{0s} - BIC_{1s}) < 9.2$, dan is de *aanwijzing* voor M_1 *sterk*;
- 4 als $(BIC_{0s} - BIC_{1s}) \geq 9.2$ dan is er *afdoende aanwijzing* voor M_1 .

Meer precies kunnen verschillen in BIC-waarden als volgt geïnterpreteerd worden. Als $(BIC_{0s} - BIC_{1s})$ gelijk is aan 5, betekent dit dat model M_0 tien keer minder waarschijnlijk is

dan model M_1 . De waarde 6 betekent dat model M_0 twintig keer minder waarschijnlijk is dan model M_1 , en de waarde 10 dat M_0 honderd keer minder waarschijnlijk is dan model M_1 .

Bij de vergelijking van twee geneste modellen zullen we de volgende beslissingsregels hanteren:

- 1 M_1 wordt verworpen ten gunste van M_0 als $(BIC_{05} < BIC_{15})$;
- 2 M_0 wordt verworpen ten gunste van M_1 als $(BIC_{05} - BIC_{15}) > 6$;
- 3 als $0 \leq (BIC_{05} - BIC_{15}) \leq 6$ zijn beide modellen gelijkwaardig.

Gezien de eerder genoemde interpretatie van het verschil in BIC-waarden, accepteren we het additionele effect in M_1 als dit model minimaal 20 keer zo waarschijnlijk is als M_0 . Indien het verschil in BIC-waarde aangeeft dat beide modellen gelijkwaardig zijn zullen we, alvorens een keuze voor een van de twee modellen te maken, de t-waarde van de betreffende parameter bekijken in het complexere model M_1 . Als deze $t \geq \sqrt{(\ln(N)+5)}$ dan zullen we deze parameter handhaven en dus de voorkeur geven aan model M_1 . In dat geval is er namelijk een sterke aanwijzing voor de betreffende parameter (zie Raftery, 1993).

4.2 De relatie tussen AVI en EMT

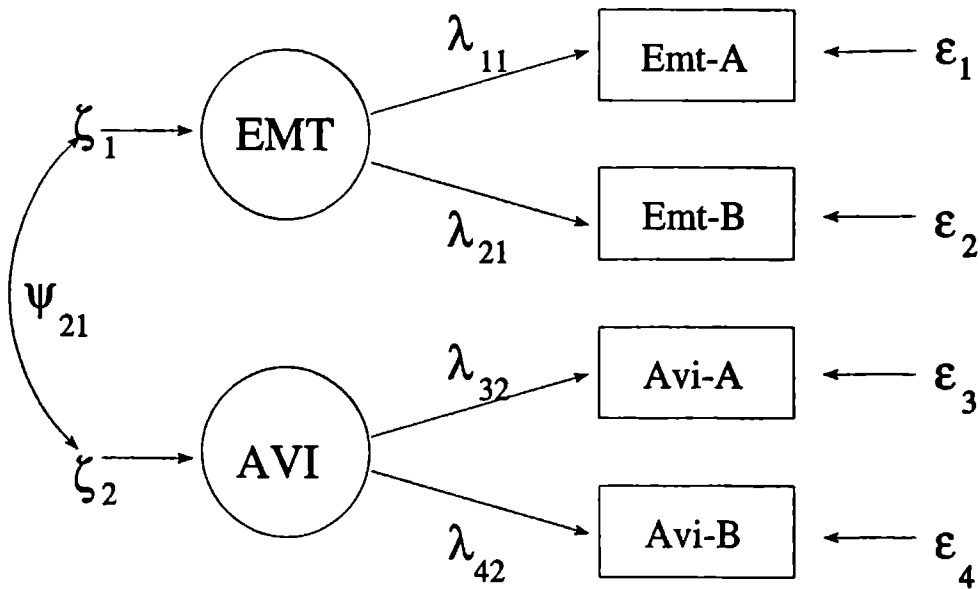
In deze paragraaf staan de volgende onderzoeksvragen centraal:

- 4 Kunnen de A- en B-versie van de AVI-toets als paralleltests worden beschouwd?
- 5 Wat is de relatie tussen de AVI-toets en de Eén-Minuut-Test (EMT) van Brus en Vooten (1973)?

Voor deze twee onderzoeksvragen hebben we één basismodel opgesteld (paragraaf 4.2.1). In paragraaf 4.2.2 presenteren we de resultaten ten aanzien van de paralleliteit van de A- en B-versie (onderzoeksvraag 4) en in paragraaf 4.2.3 ten aanzien van de relatie tussen de AVI-toets en de EMT (onderzoeksvraag 5).

4.2.1 Het basismodel

In figuur 4.1 hebben we het basismodel A weergegeven. In dit basismodel (A) veronderstellen we twee factoren (latente variabelen): EMT en AVI. We veronderstellen dat deze twee factoren met elkaar gecorreleerd zijn; de betreffende parameter is vrijgelaten. Beide factoren worden gemeten met behulp van de A- en B-versie van respectievelijk de Eén-Minuut-Test en de AVI-toetskaarten (de geobserveerde variabelen). De meeteenheid van de twee latente variabelen is vastgezet door de variantie van de disturbance-variabelen gelijk te stellen aan 1. In het basismodel A gaan we ervan uit dat de twee versies van de EMT respectievelijk de twee versies van de AVI-toetskaarten *congenerieke* tests zijn: tests met perfect gecorreleerde ware scores. Basismodel A hebben we toegepast op negen datasets, namelijk per cohort per periode.



Figuur 4.1: Basismodel A voor de relatie tussen AVI en EMT

Uitbreiding ten behoeve van onderzoeksvraag 4

We willen vervolgens toetsen of de versies van de AVI-toetskaarten beschouwd mogen worden als parallelle tests. We zullen tevens nagaan of beide versies van de EMT als parallelle tests beschouwd mogen worden, teneinde de correlatie tussen EMT en AVI te kunnen schatten in het best passende model.

Lord en Novick (1968) geven aan dat bij parallelle tests de verwachte toetsscores en de geobserveerde score-varianties gelijk dienen te zijn. Naast parallelle tests en congenerieke tests, onderscheiden ze nog tau-equivalente toetsen. Daarvoor geldt dat de ware scores gelijk zijn, maar de error varianties (mogelijk) verschillen. Bij 'essentially' tau-equivalente tests is het tevens mogelijk dat de ware scores met een additieve constante van elkaar verschillen.

Om te kunnen toetsen of de twee versies van de AVI-toets respectievelijk van de EMT aan deze voorwaarden voldoen (onderzoeksvraag 4), hebben we aan het basismodel A een aantal restricties toegevoegd.

De eerste restrictie betreft de eis voor tau-equivalente toetsen dat de ware score varianties gelijk zijn. Deze voorwaarde is in het basismodel opgenomen door gelijke factorloadingen (de λ 's in figuur 4.1) te veronderstellen voor beide versies van de AVI-toets op de factor AVI en voor beide versies van de EMT op de factor EMT (zie ook Jöreskog, 1974; Jöreskog & Sörbom, 1993a). Dit model noemen we in het vervolg model A1.

De tweede restrictie betreft de eis voor parallelle tests dat, naast gelijke ware score varianties ook sprake is van gelijke error varianties. Deze voorwaarde is opgenomen in model A1 te veronderstellen dat de errorvarianties van beide versies van eenzelfde toets

gelijk zijn (model A2). De modellen A, A1 en A2 zijn per cohort per periode toegepast. Daarbij moeten we nog opmerken dat in de tweede periode (najaar 1992) alleen de A-versie van de AVI-toetskaarten is afgenomen, waardoor we voor die periode alleen de parallelliteit van de twee versies van de EMT kunnen toetsen.

Voor volledige parallelliteit is het tevens een vereiste dat de gemiddelden op beide versies van de toets gelijk zijn. Deze restrictie hebben we, ter beperking van het aantal analyses, niet met behulp van alle negen data-sets getoetst. Voor twee perioden in de basisschool beschikken we over de gegevens van twee cohorten. Voor het voorjaar van groep 4 beschikken we over gegevens van de cohorten 3 en 4, voor het voorjaar van groep 5 over gegevens van de cohorten 4 en 5. Voor deze twee perioden hebben we modellen vergelijkbaar met de modellen A, A1 en A2 simultaan getoetst voor telkens twee cohorten. Zo toetsen we op een andere manier of sprake is van gelijke factorladingen en errorvarianties. In de simultane analyses hebben we tevens de laatste restrictie voor volledige parallelliteit getoetst, namelijk of de gemiddelden van beide versies van de AVI-toetskaarten respectievelijk de EMT gelijk zijn. Met deze simultane Lisrel-analyses hebben we eveneens getoetst of de betreffende cohorten van elkaar verschillen. Voor de simultane analyses hebben we de volgende modellen getoetst. Ten eerste model A4, dat gelijk is aan het eerder genoemde basismodel A. De meeteenheid van de latente factoren hebben we vastgelegd via de A-versie van de EMT respectievelijk de AVI-toetskaarten.

Aan model A4 hebben we achtereenvolgens de volgende restricties toegevoegd:

- de factorladingen van de geobserveerde variabelen op de latente factoren zijn gelijk in beide cohorten (model A5);
- de twee versies van de EMT respectievelijk de AVI-toetskaarten zijn tau-equivalent: gelijke factorladingen op de factor EMT respectievelijk de factor AVI (model A6);
- de twee versies van EMT respectievelijk AVI zijn, afgezien van de gemiddelden, parallel: naast gelijke factorladingen ook gelijke errorvarianties van de twee versies van dezelfde toets (model A7).

Vervolgens hebben we het model uitgebreid met gemiddelden voor de geobserveerde en de latente variabelen. Het eerste model A8 is een eenvoudige uitbreiding van het eerder genoemde model A7 met gemiddelden voor zowel de geobserveerde als de latente variabelen. In dit model veronderstellen we dat de gemiddelden op de twee latente variabelen AVI en EMT verschillen tussen de twee groepen. Ten behoeve van de identificatie van het model veronderstellen we dat de gemiddelden op de geobserveerde variabelen gelijk zijn in beide groepen. Aan model A8 hebben we vervolgens de volgende restricties opgelegd:

- de gemiddelden op de latente variabelen zijn gelijk in beide groepen (model A9);
- de gemiddelden op beide versies van de AVI-toets zijn gelijk, evenals de gemiddelden op beide versies van de EMT (model A10). Deze restrictie komt overeen met de laatste nog te toetsen eis voor parallelliteit.

Afhankelijk van de resultaten van de modeltoetsing zijn bovengenoemde restricties vervolgens voor alleen AVI of alleen EMT toegepast.

Uitbreiding ten behoeve van onderzoeksvraag 5

Ter beantwoording van de vijfde onderzoeksvraag willen we toetsen of de EMT en de AVI-toetskaarten hetzelfde meten. Beide toetsen zijn immers bedoeld voor het meten van technisch lezen en in beide speelt de snelheid van hardop lezen een belangrijke rol. Het belangrijkste verschil is het toetsmateriaal: in de AVI-toetskaarten wordt gebruik gemaakt van teksten, in de EMT van losse woorden.

Aan model A2 hebben we daarvoor nog één restrictie toegevoegd, namelijk dat de correlatie tussen de twee factoren (AVI en EMT) gelijk is aan 1. Het toevoegen van deze restrictie komt overeen met het veronderstellen van één latente variabele voor de twee versies van de AVI-toetskaarten en de twee versies van de EMT. In het resulterende model A3 hebben we deze restrictie opgenomen door de correlatie tussen de residuen (disturbance variabelen) van beide latente factoren gelijk aan 1 te stellen. Aangezien een correlatie van 1 wel een erg streng criterium is, hebben we tevens getoetst of deze correlatie gelijk gesteld mag worden aan 0,95 (model A3a).

4.2.2 Parallelliteit van A- en B-versies van AVI en EMT

We beschikken, zoals eerder gezegd, over de gegevens van drie cohorten over drie perioden. We hebben de modellen A tot en met A3a per cohort per periode toegepast. In tabel 4.8 presenteren we de fit-maten behorend bij de verschillende modellen.

Tabel 4.8: Modeltoetsing betreffende de parallelliteit van de versie A en B en de relatie tussen AVI en EMT, per cohort per periode

	χ^2	df	p	RMSEA	$P_{close\ fit}$	GFI	CFI	BIC
COHORT 3 VOORJAAR 1992 (N=954)								
A Basismodel, 2 factoren	0,300	1	0,584	0,000	0,859	1,000	1,000	-7,947
A1=A + AVI + EMT tau-eq.	2,194	3	0,533	0,000	0,957	0,999	1,000	-22,547
A2=A + AVI + EMT parallel	5,877	5	0,318	0,014	0,958	0,997	1,000	-35,358
A3=A2 + $\psi_{emt-avi} = 1$	599,632	6	0,000	0,322	0,000	0,798	0,880	550,150
A3a=A2 + $\psi_{emt-avi} = 0,95$	198,494	6	0,000	0,183	0,000	0,896	0,961	149,012
NAJAAR 1992 (N=949)								
A Basismodel, 2 factoren	0,000	0	1,000					0,000
A1=A + EMT tau-eq.	7,198	1	0,007	0,081	0,126	0,995	0,998	-0,756
A2=A + EMT parallel	11,204	2	0,004	0,070	0,163	0,992	0,998	-4,704
A3=A2 + $\psi_{emt-avi} = 1$	1403,758	3	0,000	0,702	0,000	0,660	0,629	1379,896
A3a=A2 + $\psi_{emt-avi} = 0,95$	254,662	3	0,000	0,297	0,000	0,859	0,933	230,800
VOORJAAR 1993 (N=944)								
A Basismodel, 2 factoren	1,323	1	0,250	0,019	0,654	0,999	1,000	-6,913
A1=A + AVI + EMT tau-eq.	14,960	3	0,002	0,065	0,186	0,992	0,998	-9,749
A2=A + AVI + EMT parallel	18,739	5	0,002	0,054	0,355	0,990	0,997	-22,443
A3=A2 + $\psi_{emt-avi} = 1$	338,373	6	0,000	0,242	0,000	0,862	0,932	288,954
A3a=A2 + $\psi_{emt-avi} = 0,95$	93,525	6	0,000	0,124	0,000	0,951	0,982	44,106

	χ^2	df	p	RMSEA	$p_{close\ fit}$	GFI	CFI	BIC
COHORT 4 VOORJAAR 1992 (N=937)								
A Basismodel, 2 factoren	2,100	1	0,147	0,034	0,534	0,999	1,000	-6,129
A1=A + AVI + EMT tau-eq.	11,533	3	0,009	0,055	0,342	0,994	0,998	-13,154
A2=A + AVI + EMT parallel	12,418	5	0,030	0,040	0,684	0,994	0,999	-28,727
A3=A2 + $\psi_{emt-avi} = 1$	187,432	6	0,000	0,180	0,000	0,909	0,968	138,058
A3a=A2 + $\psi_{emt-avi} = 0,95$	14,170	6	0,028	0,038	0,744	0,993	0,999	-35,204
NAJAAR 1992 (N=846)								
A Basismodel, 2 factoren	0,000	0	1,000					0,000
A1=A + EMT tau-eq.	0,043	1	0,836	0,000	0,942	1,000	1,000	-7,796
A2=A + EMT parallel	0,406	2	0,816	0,000	0,973	1,000	1,000	-15,272
A3=A2 + $\psi_{emt-avi} = 1$	842,128	3	0,000	0,575	0,000	0,704	0,721	818,611
A3a=A2 + $\psi_{emt-avi} = 0,95$	161,517	3	0,000	0,250	0,000	0,888	0,947	138,000
VOORJAAR 1993 (N=843)								
A Basismodel, 2 factoren	0,015	1	0,903	0,000	0,966	1,000	1,000	-8,108
A1=A + AVI + EMT tau-eq.	6,324	3	0,097	0,036	0,658	0,996	0,999	-18,046
A2=A + AVI + EMT parallel	11,659	5	0,040	0,040	0,670	0,993	0,998	-28,957
A3=A2 + $\psi_{emt-avi} = 1$	237,102	6	0,000	0,214	0,000	0,883	0,943	188,362
A3a=A2 + $\psi_{emt-avi} = 0,95$	50,097	6	0,000	0,093	0,001	0,970	0,989	1,357
COHORT 5 VOORJAAR 1992 (N=852)								
A Basismodel, 2 factoren	1,194	1	0,275	0,015	0,648	0,999	1,000	-6,940
A1=A + AVI + EMT tau-eq.	3,166	3	0,367	0,008	0,892	0,998	1,000	-21,236
A2=A + AVI + EMT parallel	9,153	5	0,103	0,031	0,811	0,995	0,999	-31,516
A3=A2 + $\psi_{emt-avi} = 1$	325,762	6	0,000	0,250	0,000	0,848	0,923	276,959
A3a=A2 + $\psi_{emt-avi} = 0,95$	66,533	6	0,000	0,109	0,000	0,959	0,985	17,730
NAJAAR 1992 (N=806)								
A Basismodel, 2 factoren	0,000	0	1,000					0,000
A1=A + EMT tau-eq.	0,066	1	0,797	0,000	0,924	1,000	1,000	-7,724
A2=A + EMT parallel	0,225	2	0,894	0,000	0,984	1,000	1,000	-15,356
A3=A2 + $\psi_{emt-avi} = 1$	891,368	3	0,000	0,607	0,000	0,691	0,626	867,996
A3a=A2 + $\psi_{emt-avi} = 0,95$	341,580	3	0,000	0,374	0,000	0,794	0,858	318,208
VOORJAAR 1993 (N=802)								
A Basismodel, 2 factoren	0,034	1	0,855	0,000	0,946	1,000	1,000	-8,042
A1=A + AVI + EMT tau-eq.	2,064	3	0,559	0,000	0,942	0,999	1,000	-22,164
A2=A + AVI + EMT parallel	10,145	5	0,071	0,036	0,732	0,994	0,999	-30,234
A3=A2 + $\psi_{emt-avi} = 1$	426,866	6	0,000	0,296	0,000	0,817	0,880	378,411
A3a=A2 + $\psi_{emt-avi} = 0,95$	196,006	6	0,000	0,199	0,000	0,877	0,946	147,551

In kolom 1 staat een korte omschrijving van het model, de kolommen 2 tot en met 4 bevatten de χ^2 met het bijbehorende aantal vrijheidsgraden (df) en de p-waarde, in de kolommen 5 en 6 staan de RMSEA en de bijbehorende p-waarde voor 'close fit' (toetsing of de RMSEA significant groter is dan 0,05) en de kolommen 7 tot en met 9 bevatten achtereenvolgens de Goodness-of-fit index (GFI), de Comparative fit index (CFI) en de BIC.

In het basismodel A veronderstellen we twee aparte factoren voor de AVI-toetskaarten en de EMT. In de tweede periode (najaar 1992) kan dit model niet getoetst worden, aangezien het in deze periode het verzadigde model is en dus altijd volledig bij de data past. Wanneer we kijken in tabel 4.8 naar cohort 3 voorjaar 1992 (ofwel de eerste periode) dan past het model bij de data. Vergelijkbare resultaten vinden we voor de derde periode van cohort 3 en de eerste en derde periode van de cohorten 4 en 5. In alle gevallen past het basismodel, met aparte factoren voor de AVI-toets en de EMT, waarbij de twee versies van een toets als congeneriek worden beschouwd (tests met perfect gecorreleerde ware scores), bij de data.

Model A1 verschilt van model A dat daarin verondersteld wordt dat de twee versies van de AVI-toets, net als de twee versies van de EMT, tau-equivalent zijn. Dat wil zeggen dat de factorlading van versie A van de AVI-toets op de (latente) AVI-factor gelijk is aan de factorlading van versie B en dat de factorlading van EMT versie A op de EMT-factor gelijk is aan de factorlading van EMT versie B.

Wanneer we opnieuw kijken naar de eerste periode van cohort 3 dan zien we dat model A1 eveneens past bij de data. De resultaten voor de tweede periode en de derde periode van cohort 3 wijzen er ook op dat het model bij de data past. In beide gevallen geeft de toetsing van de χ^2 aan dat het model significant afwijkt van de data, maar geven de overige maten voor overall fit aan dat het model wel bij de data past (lage RSMEA, hoge GFI en CFI en een negatieve BIC). Een vergelijkbaar resultaat vinden we voor cohort 4 in de eerste periode. Voor de overige perioden van dit cohort en voor alle perioden van cohort 5 geeft de modeltoetsing aan dat model A1 bij de data past.

Het laatste model dat belangrijk is voor de beantwoording van de vierde onderzoeksvraag is model A2, waarin we paralleliteit (afgezien van de gemiddelden) van de versies van de AVI-toets en de EMT veronderstellen. Naast gelijke factorladingen van de twee versies van een toets op de factor veronderstellen we daarbij tevens dat de errorvarianties van de versies gelijk zijn. Oftewel: de meetfout van versie A van de AVI-toets is gelijk aan die van versie B. Hetzelfde geldt voor de twee versies van de EMT.

De fit-maten behorende bij model A3 in de eerste periode van cohort 3 geven aan dat ook dit model bij de data past. Evenals bij model A2 geeft in een aantal gevallen de toetsing van de χ^2 aan dat het model niet bij de data past terwijl de overige fit-maten het omgekeerde aangeven (cohort 3 periode 2 en 3 en cohort 4 periode 1 en 3). In alle andere gevallen wijzen alle fit-maten erop dat model A2 bij de data past.

Op grond van de resultaten concluderen we dat de veronderstelling dat de twee versies van de AVI-toets en de twee versies van de EMT, afgezien van de gemiddelden, parallel zijn, niet strijdig is met de data: model A2 past immers bij de data. Maar ook de model-

len A en A2 passen bij de data. Op basis van vergelijking van de modelresultaten dienen we tot een keuze te komen.

De toetsing van het verschil in χ^2 tussen de modellen A2 en A geeft voor de eerste periode van cohort 3 aan dat model A2 de voorkeur heeft ($\chi^2_{A2-A}=5,6$, $df=4$, $p>0,20$). De hypothese dat de versies van de AVI en de EMT parallel zijn kan niet verworpen worden. Voor de overige twee perioden van cohort 3 is het verschil in χ^2 tussen model A2 en A zo groot dat deze hypothese wel verworpen zou moeten worden (periode 2: $\chi^2_{A2-A}=11,2$, $df=2$, $p<0,01$; periode 3: $\chi^2_{A2-A}=17,4$, $df=4$, $p<0,01$). Hetzelfde geldt voor de vergelijking tussen de model A1 en A.

Voor cohort 4 geeft de toetsing van het verschil in χ^2 tussen model A2 en A aan dat in periode 1 ($\chi^2_{A2-A}=10,3$, $df=4$, $p<0,05$) en in periode 3 ($\chi^2_{A2-A}=11,6$, $df=4$, $p<0,05$) de hypothese van paralleliteit verworpen moet worden. Voor periode 3 kan wel de hypothese van tau-equivalentie gehandhaafd blijven ($\chi^2_{A1-A}=6,3$, $df=4$, $p>0,10$). Voor periode 2 van cohort 3 kan volgens deze methode de hypothese van volledige paralleliteit gehandhaafd blijven ($\chi^2_{A2-A}=0,4$, $df=2$, $p>0,20$).

De resultaten van de toetsing van verschillen in χ^2 voor cohort 5 geven aan dat in de perioden 1 en 2 de hypothese van volledige paralleliteit gehandhaafd kan blijven, terwijl in periode 3 de hypothese van tau-equivalentie gehandhaafd blijft.

De BIC van model A2, het meest restrictieve van de drie relevante modellen, is in de eerste periode van cohort 3 duidelijk lager dan die van de modellen A en A1 (-35,358 versus -7,947 en -22,547). Op basis hiervan kunnen we de hypothese van paralleliteit, afgezien van de gemiddelden, handhaven voor deze data-set. Dit resultaat komt overeen met dat van de toetsing van de verschillen in χ^2 . Ook in de andere twee perioden van cohort 3 is de BIC van model A2 het laagst, zodat we op basis daarvan ook de hypothese van paralleliteit zouden kunnen handhaven. Deze resultaten komen niet overeen met de eerder gepresenteerde toetsing van verschillen in χ^2 .

Model A2 is in de cohorten 4 en 5 in alle drie perioden eveneens het model met de laagste BIC, zodat we op basis van dit criterium de hypothese van paralleliteit kunnen handhaven.

Wat kunnen we nu concluderen ten aanzien van de paralleliteit van de twee versies van de AVI-toetskaarten en de twee versies van de EMT? Op basis van de eerder besproken resultaten kunnen we concluderen dat zowel voor de twee versies van de AVI-toetskaarten als voor de twee versies van de EMT sprake is van paralleliteit, afgezien van mogelijke verschillen in gemiddelden. Van de vijf fit-maten wijst alleen de χ^2 -waarde er in een aantal gevallen op dat deze hypothese verworpen moet worden. De RMSEA, de GFI en de CFI geven in alle gevallen aan dat het betreffende model A2 bij de data past en op basis van de vergelijking van de BIC's kunnen we afleiden dat model A2 de voorkeur geniet.

Zoals in paragraaf 4.2.1 reeds is gezegd, zullen we voor twee perioden, namelijk het voorjaar van groep 4 en 5, simultaan voor twee cohorten Lisrel-modellen toetsen. Daarbij zal tevens de laatste eis voor paralleliteit (gelijke gemiddelden) worden getoetst. In tabel 4.9 zijn de resultaten van deze analyses weergegeven.

Tabel 4.9: Toetsing modellen voor AVI en EMT voor twee groepen tegelijk voor het voorjaar van groep 4 groep 5

	χ^2	df	p	RMSEA	Pclose fit	GFI	CFI	BIC
VOORJAAR GROEP 4 (N ₁ =944, N ₂ =937)								
A4 Basismodel	3,423	2	0,181	0,020	0,924	0,999	1,000	-14,429
A5=A4+ factorladingen gelijk in de cohorten	4,316	4	0,365	0,006	0,997	0,999	1,000	-31,387
A6=A5 + EMT+AVI tau-eq	26,494	6	0,000	0,043	0,740	0,994	0,998	-27,061
A7=A5 + EMT+AVI parallel	31,157	10	0,001	0,034	0,977	0,994	0,998	-58,102
A8=A7 + gemiddelden	46,797	12	0,000	0,039	0,925	0,993	0,997	-60,313
A9=A8 + gemiddelde latente variabelen gelijk	71,608	14	0,000	0,047	0,668	0,994	0,995	-53,354
A9a=A8 + gemiddelde EMT gelijk	53,433	13	0,000	0,041	0,903	0,993	0,996	-62,603
A9b=A8 + gemiddelde AVI gelijk	64,891	13	0,000	0,046	0,697	0,993	0,995	-51,145
A10=A9a + gelijke intercepten versie A en B	220,498	15	0,000	0,085	0,000	0,992	0,981	86,610
A10a=A9a + gelijke intercepten EMT-versies	215,995	14	0,000	0,088	0,000	0,993	0,981	91,033
A10b=A9a + gelijke intercepten AVI-versies	57,936	14	0,000	0,041	0,908	0,993	0,996	-67,026
VOORJAAR GROEP 5 (N ₁ =843, N ₂ =852)								
A4 Basismodel	1,209	2	0,546	0,000	0,982	0,999	1,000	-16,434
A5=A4+ factorladingen gelijk in de cohorten	3,769	4	0,438	0,000	0,997	0,999	1,000	-31,518
A6=A5 + EMT+AVI tau-eq	9,489	6	0,148	0,019	0,995	0,998	1,000	-43,441
A7=A5 + EMT+AVI parallel	20,812	10	0,022	0,025	0,997	0,995	0,999	-67,405
A8=A7 + gemiddelden	31,840	12	0,001	0,031	0,991	0,995	0,998	-74,021
A9=A8 + gemiddelde latente variabelen gelijk	32,208	14	0,004	0,028	0,999	0,995	0,998	-91,296
A9a=A8 + gemiddelde EMT gelijk	32,197	13	0,002	0,030	0,996	0,995	0,998	-82,486
A9b=A8 + gemiddelde AVI gelijk	32,159	13	0,002	0,030	0,996	0,995	0,998	-82,524
A10=A9 + gelijke intercepten versie A en B	84,543	16	0,000	0,050	0,459	0,995	0,992	-56,605
A10a=A9 + gelijke intercepten EMT-versies	83,681	15	0,000	0,052	0,360	0,995	0,992	-48,645
A10b=A9 + gelijke intercepten AVI-versies	33,069	15	0,005	0,027	0,999	0,995	0,998	-99,257

Wanneer we naar de resultaten van het voorjaar van groep 4 kijken, zien we het volgende. Het basismodel A4 past goed bij de data, evenals model A5 waarin we veronderstellen dat de factorstructuur in beide groepen gelijk is.

De restrictie dat de twee versies van de AVI-toetskaarten en de twee versies van de EMT tau-equivalent zijn, leidt ertoe dat het model (A6) volgens de χ^2 niet meer bij de data past. De waarden van de RMSEA, de GFI, de CFI en de BIC daarentegen wijzen erop dat het model wel bij de data past. De BIC van model A6 is wel iets hoger dan die van model A5, maar dit weegt op tegen de winst van twee vrijheidsgraden.

Voor model A7, waarin we, afgezien van de gemiddelden, paralleliteit van de versies veronderstellen, zijn de resultaten vergelijkbaar: de χ^2 -waarde geeft aan dat het model niet bij de data past, de overige fit-maten geven dat wel aan. De BIC van dit model is het laagst van de modellen A4 tot en met A7 en de RMSEA is goed.

Kortom, de resultaten van deze analyses wijzen op hetzelfde als de eerder gepresenteerde analyses per cohort per periode: de hypothese van paralleliteit (afgezien van eventuele verschillen in gemiddelden) kan gehandhaafd blijven. Een extra uitkomst van deze analyses is dat de factorstructuur in de twee onderzochte cohorten gelijk is.

Model A8 is het eerste model waarin de gemiddelden van de geobserveerde en latente variabelen zijn opgenomen. Net als in model A7 veronderstellen we daarin paralleliteit van de versies, afgezien van een eventueel verschil in gemiddelden. De restrictie dat de gemiddelden op de latente variabelen gelijk zijn in beide cohorten (model A9) leidt tot een grote toename in χ^2 en een stijging van de BIC. We hebben vervolgens in de modellen A9a en A9b alleen het gemiddelde van de factor EMT respectievelijk de factor AVI gelijk gesteld in beide cohorten. De stijging in χ^2 en BIC kunnen vooral toegeschreven worden aan het gelijkstellen van het gemiddelde van de AVI-factor in beide cohorten. Immers, wanneer we bijvoorbeeld naar de BIC kijken, zien we dat het gelijkstellen van het gemiddelde van de EMT-factor in beide cohorten leidt tot een daling (model A9a versus model A8). Het toevoegen aan model A9a van de restrictie dat ook het gemiddelde op de AVI-factor gelijk is in de twee groepen (dit is model A9) leidt tot een forse stijging van de BIC.

Tot slot hebben we model A9a nog uitgebreid met de restricties dat de gemiddelden op de A- en B-versie van de AVI-toets aan elkaar gelijk zijn en dat de gemiddelden op de twee versies van de EMT-toets aan elkaar gelijk zijn. Dit komt overeen met de laatste nog te toetsen eis voor paralleliteit. Uit de modellen A10a tot en met A10b blijkt dat deze restrictie voor de twee versies van EMT ertoe leidt dat de modellen niet bij de data passen (A10 en A10a). Model A10b, waarin alleen wordt verondersteld dat de gemiddelden op de twee versies van de AVI-toets gelijk zijn, past daarentegen wel bij de data.

Voor maart groep 5 hebben we vergelijkbare analyses uitgevoerd, waarbij we niet zo uitgebreid willen stilstaan. De resultaten komen vrijwel geheel overeen met die voor maart groep 4, met één verschil: het gemiddelde op de AVI-factor is eveneens gelijk in beide cohorten.

Conclusies

Op basis van de analyses waarbij we telkens twee groepen konden vergelijken, kunnen we het volgende concluderen.

Ten eerste kunnen we concluderen dat de twee versies van de AVI-toetskaarten volledig parallel zijn. Voor de twee versies van de EMT geldt dat deze, afgezien van het verschil in gemiddelden, parallel zijn.

4.2.3 De samenhang tussen AVI en EMT

Voor de beantwoording van de vijfde onderzoeksvraag hebben we de modellen A3 en A3a getoetst, gebruikmakend van de databestanden per cohort per periode. In tabel 4.8 zijn de toetsingsresultaten voor deze modellen weergegeven.

Uit de tabel blijkt dat model A3, waarin we in feite één factor veronderstellen voor de AVI-toetskaarten en de EMT, niet bij de data past, voor geen enkel cohort en geen enkele periode. De χ^2 -waarden zijn zeer hoog (tussen 187,432 en 1403,758), de RMSEA is duidelijk hoger dan 0,05 (tussen 0,180 en 0,702), de GFI en de CFI zijn meestal lager dan 0,900 en de BIC is groter dan 0 (tussen 138,058 en 1379,896). Hieruit mogen we concluderen dat de correlatie tussen de EMT en de AVI niet gelijk is aan 1: beide toetsen meten niet precies hetzelfde.

Model A3a, waarin we een correlatie van 0,95 veronderstellen tussen beide latente variabelen, past alleen bij de data van het voorjaar 1992 van cohort 4. Voor deze dataset geldt dat dit model ook beter bij de data past dan model A2, waarin de correlatie tussen beide latente variabelen is vrijgelaten.

In tabel 4.10 hebben we de met model A2 geschatte correlaties tussen de latente variabelen EMT en AVI weergegeven per cohort per periode in het basisonderwijs.

Tabel 4.10: Correlaties tussen de factoren EMT en AVI (model A2) per cohort per periode in het basisonderwijs

	periode in de basisschoolperiode					
	voorjaar groep 3	najaar groep 4	voorjaar groep 4	najaar groep 5	voorjaar groep 5	najaar groep 6
Cohort 3	0,855	0,861	0,892			
Cohort 4			0,956	0,875	0,906	
Cohort 5					0,899	0,807
						0,830

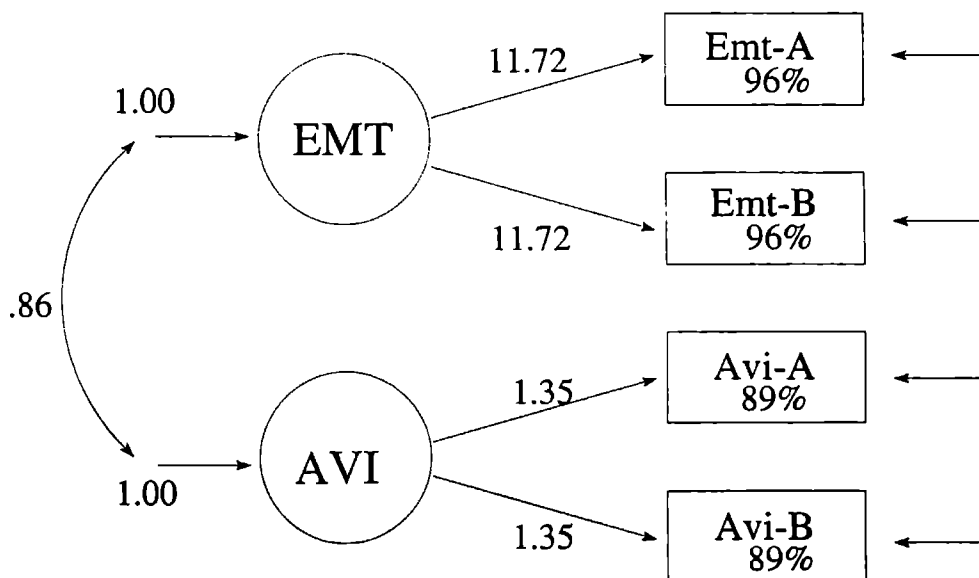
Gemiddeld genomen ligt de voor onbetrouwbaarheid gecorrigeerde correlatie tussen AVI en EMT tussen 0,85 en 0,90. Een uitschieter naar boven vinden we in het voorjaar van groep 4 in cohort 4 (0,96), uitschieters naar beneden zien we in groep 6 van cohort 5. De correlatie lijkt redelijk stabiel in de groepen 4 en 5, maar neemt iets af in groep 6. Op zoek naar een mogelijke oorzaak hiervoor hebben we spreidingsdiagrammen gemaakt van de gemiddelde scores op de twee versies van de AVI-toets met de gemiddelde scores op de twee versies van de EMT (zie bijlage 13).

In het voorjaar van groep 3 is bij de AVI-toets sprake van een bodemeffect: een groot deel van de leerlingen beheerst niveau 1 nog niet (score 0; zie ook tabel 3.14 in hoofdstuk 3). In oktober en maart van groep 6 is duidelijk sprake van een plafondeffect bij de AVI-toetskaarten, terwijl dat bij de Eén-Minut-Test zeker nog niet het geval is. Dit plafondeffect bij de AVI-toetskaarten kan een oorzaak zijn van de lagere correlaties tussen de factoren AVI en EMT die we in deze perioden hebben gevonden.

Conclusie

Samenvattend kunnen we als antwoord op de vierde onderzoeksvraag vaststellen dat de twee versies van de AVI-toetskaarten volledig parallel zijn. Dat wil zeggen dat de factorladingen op de latente variabele, de errorvarianties en de gemiddelden van de A- en B-versie van de AVI-toetskaarten gelijk zijn. Voor de twee versies van de Eén-Minut-Test geldt dat wel de factorladingen en de errorvarianties maar niet de gemiddelden gelijk zijn.

Ter beantwoording van de vijfde onderzoeksvraag kunnen we concluderen dat de correlatie tussen de twee latente variabelen EMT en AVI wel hoog, maar niet gelijk aan 1 is. De correlatie tussen beide factoren ligt voor groep 3, 4 en 5 tussen de 0,85 en 0,90 en is voor groep 6 iets lager (0,81-0,83), wellicht als gevolg van het plafondeffect bij de AVI-toetskaarten. Dit betekent dat de AVI-toetskaarten en de EMT niet precies hetzelfde aspect van de technische leesvaardigheid meten.



Figuur 4.2: Ongestandaardiseerde parameterschattingen van model A2 voor cohort 3 periode 1

Als uitgangspunt voor de overige analyses in dit hoofdstuk zullen we het model A2 nemen, waarin beide versies van de EMT en de AVI als parallel worden beschouwd (met uitzondering van eventuele verschillen in gemiddelden) en we een aparte factor voor AVI en EMT onderscheiden. Ter illustratie geven we in figuur 4.2 de ongestandaardiseerde parameterschattingen voor model A2 weer gebaseerd op de gegevens van het voorjaar 1992 van cohort 3. In bijlage 14 staan voor alle negen datasets de ongestandaardiseerde parameterschattingen voor model A2. Bijlage 15 bevat de ongestandaardiseerde parameterschattingen uit de simultane Lisrel-analyses voor het voorjaar van groep 4 en groep 5. In figuur 4.2 is zichtbaar dat de factorladingen van beide versies van de AVI-toetskaarten op de AVI-factor aan elkaar gelijk zijn, evenals de factorladingen van beide versies van de Eén-Minuuut-Test op de EMT-factor. De correlatie tussen beide latente variabelen bedraagt 0,86. Tevens zien we dat de EMT-factor 96% van de variantie van beide versies van de Eén-Minuuut-Test verklaart. De AVI-factor verklaart 89% van de variantie van beide versies van de AVI-toetskaarten. De percentages verklaarde variantie, die beschouwd mogen worden als de betrouwbaarheid van de toetsen, zijn redelijk stabiel in de negen datasets: tussen 93% en 96% voor de Eén-Minuuut-Test en tussen 85% en 91% voor de AVI-toetskaarten.

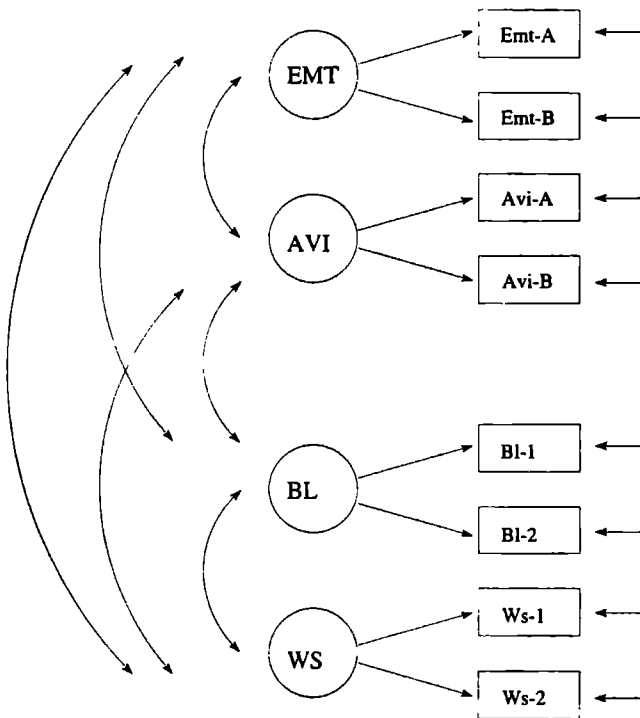
4.3 De relatie met begrijpend lezen en woordenschat per cohort per periode

In deze paragraaf staat de zesde onderzoeksvraag centraal: wat is de relatie tussen de AVI-toets en toetsen voor begrijpend lezen en woordenschat? In deze paragraaf zullen we deze vraag beantwoorden met behulp van modellen per cohort per periode. In paragraaf 4.4 bespreken we longitudinale modellen betreffende de relatie tussen de AVI-toetskaarten en toetsen voor begrijpend lezen en woordenschat.

4.3.1 Het basismodel

Om de zesde onderzoeksvraag te kunnen beantwoorden, hebben we model A2 uitgebreid met twee extra factoren: begrijpend lezen (BL) en woordenschat (WS). Beide factoren worden gemeten met twee geobserveerde variabelen. De meeteenheid van de latente factoren hebben we vastgelegd door de varianties van de latente variabelen gelijk te stellen aan 1. Tussen de vier factoren (AVI, EMT, BL en WS) veronderstellen we correlaties. In figuur 4.3 hebben we dit model B weergegeven.

In hoofdstuk 1 hebben we kort gesproken over de relatie tussen begrijpend lezen en woordenschat. Vanuit de verbale bekwaamheidsvisie zouden toetsen voor begrijpend lezen en woordenschat metingen zijn van hetzelfde onderliggende concept 'verbale bekwaamheid' (zie paragraaf 1.4.2). Om het idee te toetsen dat aan de toetsen voor begrijpend lezen en woordenschat één enkele factor ten grondslag ligt hebben we aan model B de restrictie toegevoegd dat de correlatie tussen de factoren Begrijpend Lezen en Woordenschat gelijk is aan 1 (model B1). Omdat een correlatie van 1 een streng criterium is hebben we tevens getoetst of de correlatie gelijk gesteld mag worden aan 0,95 (model B1a).



Figuur 4.3: Basismodel B voor de relatie met Begrijpend Lezen en Woordenschat per periode

4.3.2 De relatie tussen Begrijpend Lezen en Woordenschat

In tabel 4.11 hebben we per cohort per periode de resultaten van de toetsing van de modellen B, B1 en B1a weergegeven.

Het basismodel B past, gezien de verschillende fit-maten, bij de data. Daarentegen zien we dat in veel gevallen de toetsing van de χ^2 aangeeft dat het model niet bij de data past, met uitzondering van de derde periode van cohort 4 en de eerste en derde periode van cohort 5.

De resultaten voor model B1 (de correlatie tussen Begrijpend Lezen en Woordenschat is gelijk aan 1) zijn minder eenduidig. In de eerste periode van cohort 3 past dit model bij de data volgens alle fit-maten behalve de χ^2 . De BIC van dit meer restrictieve model is lager dan de BIC van model B, hetgeen erop wijst dat we één factor voor Begrijpend Lezen en Woordenschat mogen veronderstellen. Model B1 zou ook in de tweede periode de voorkeur genieten boven model B, zowel gezien de BIC (-74,833 versus -68,883) als gezien het verschil in χ^2 ($\chi^2_{B1-B}=2,9$, $df=1$, $p>0,05$). In de derde periode daarentegen zou model B de voorkeur genieten boven model B1. Zowel de BIC als de χ^2 van model B1 liggen beduidend hoger dan bij model B.

Tabel 4.11: Toetsing van de modellen betreffende de relatie met Begrijpend Lezen en Woordenschat, per cohort per periode

	χ^2	df	p	RMSEA	$P_{close\ fit}$	GFI	CFI	BIC
COHORT 3 VOORJAAR 1992 (N=954)								
B=A2 + BL + WS	33,537	18	0,014	0,030	0,984	0,991	0,998	-127,385
B1=B + $\psi_{bl-ws} = 1$	40,751	19	0,003	0,035	0,957	0,990	0,997	-129,111
B1a=B + $\psi_{bl-ws} = 0,95$	33,785	19	0,020	0,029	0,991	0,991	0,998	-136,077
NAJAAR 1992 (N=949)								
B=A2 + BL + WS	27,932	11	0,003	0,040	0,785	0,992	0,998	-68,883
B1=B + $\psi_{bl-ws} = 1$	30,783	12	0,002	0,041	0,788	0,991	0,997	-74,833
B1a=B + $\psi_{bl-ws} = 0,95$	35,160	12	0,000	0,045	0,650	0,990	0,997	-70,456
VOORJAAR 1993 (N=944)								
B=A2 + BL + WS	57,265	18	0,000	0,048	0,563	0,986	0,995	-103,467
B1=B + $\psi_{bl-ws} = 1$	86,515	19	0,000	0,061	0,070	0,979	0,991	-83,147
B1a=B + $\psi_{bl-ws} = 0,95$	60,069	19	0,000	0,048	0,576	0,985	0,995	-109,593
COHORT 4 VOORJAAR 1992 (N=937)								
B=A2 + BL + WS	35,402	18	0,008	0,032	0,972	0,991	0,998	-125,196
B1=B + $\psi_{bl-ws} = 1$	77,209	19	0,000	0,057	0,172	0,980	0,993	-92,311
B1a=B + $\psi_{bl-ws} = 0,95$	44,550	19	0,001	0,038	0,911	0,989	0,997	-124,970
NAJAAR 1992 (N=846)								
B=A2 + BL + WS	29,208	11	0,002	0,044	0,655	0,990	0,997	-66,343
B1=B + $\psi_{bl-ws} = 1$	75,243	12	0,000	0,079	0,002	0,975	0,988	-28,994
B1a=B + $\psi_{bl-ws} = 0,95$	37,377	12	0,000	0,050	0,466	0,987	0,995	-66,860
VOORJAAR 1993 (N=843)								
B=A2 + BL + WS	27,688	18	0,067	0,025	0,992	0,992	0,998	-131,007
B1=B + $\psi_{bl-ws} = 1$	69,755	19	0,000	0,056	0,214	0,980	0,992	-97,757
B1a=B + $\psi_{bl-ws} = 0,95$	39,043	19	0,004	0,035	0,935	0,989	0,997	-128,469
COHORT 5 VOORJAAR 1992 (N=852)								
B=A2 + BL + WS	27,880	18	0,064	0,025	0,992	0,992	0,998	-131,007
B1=B + $\psi_{bl-ws} = 1$	28,338	19	0,077	0,024	0,995	0,992	0,998	-139,376
B1a=B + $\psi_{bl-ws} = 0,95$	30,807	19	0,042	0,027	0,990	0,991	0,998	-136,907
NAJAAR 1992 (N=806)								
B=A2 + BL + WS	31,044	11	0,001	0,048	0,545	0,989	0,995	-63,974
B1=B + $\psi_{bl-ws} = 1$	80,476	12	0,000	0,084	0,001	0,973	0,985	-23,180
B1a=B + $\psi_{bl-ws} = 0,95$	44,799	12	0,000	0,058	0,206	0,985	0,993	-58,857
VOORJAAR 1993 (N=802)								
B=A2 + BL + WS	22,004	18	0,232	0,017	0,998	0,993	0,999	-135,794
B1=B + $\psi_{bl-ws} = 1$	99,947	19	0,000	0,073	0,004	0,969	0,987	-66,617
B1a=B + $\psi_{bl-ws} = 0,95$	44,290	19	0,001	0,041	0,821	0,986	0,996	-122,274

Voor cohort 4 geven de resultaten van de modeltoetsingen aan dat model B de voorkeur geniet boven model B1. Voor alle drie perioden geldt dat de hypothese dat de correlatie tussen Begrijpend Lezen en Woordenschat gelijk is aan 1, verworpen moet worden gezien de verschillen in χ^2 en BIC tussen beide modellen. Voor de tweede periode geldt tevens dat de RMSEA van model B1 significant groter is dan 0,05.

Tot slot geldt voor cohort 5 dat in de eerste periode de betreffende hypothese dat de correlatie tussen Begrijpend Lezen en Woordenschat gelijk is aan 1, gehandhaafd kan worden. In de overige twee perioden dient deze hypothese verworpen te worden (zowel gezien de χ^2 , de BIC als de RMSEA).

Wanneer we de toetsingsresultaten van model B1a (correlatie tussen Begrijpend Lezen en Woordenschat is gelijk aan 0,95) vergelijken met die van basismodel B, kunnen we het volgende concluderen. Voor cohort 3 geldt dat in de eerste en derde periode model B1a de voorkeur geniet boven model B en ook boven model B1. Voor cohort 4 is model B1a in alle drie perioden gelijkwaardig aan model B. Tot slot zien we dat voor cohort 5 in de eerste periode model B1 de voorkeur heeft boven model B1a, in de tweede periode de modellen B en B1a gelijkwaardig zijn en in de derde periode model B de voorkeur heeft boven model B1a.

Samenvattend blijkt dat in twee gevallen (cohort 3 periode 2 en cohort 5 periode 1) de hypothese dat de correlatie tussen Begrijpend Lezen en Woordenschat gelijk is aan 1, niet verworpen kan worden. Voor bijna alle andere datasets zijn de modellen waarin de correlatie tussen Begrijpend Lezen en Woordenschat gelijk wordt gesteld aan 0,95, gelijkwaardig aan de modellen waarin de correlatie wordt vrijgelaten. Deze resultaten wijzen erop dat de factoren Begrijpend Lezen en Woordenschat zeer hoog gecorreleerd zijn, hetgeen ondersteunend is voor het idee dat beide soorten toetsen metingen zijn van het concept 'verbale bekwaamheid'. Aangezien het onderscheid tussen beide typen toetsen in de praktijk van het onderwijs gebruikelijk is, kiezen we er echter voor om het onderscheid in Begrijpend lezen en Woordenschat te handhaven. In paragraaf 4.3.3 komen we nog terug op deze relatie.

4.3.3 De relatie tussen AVI en Begrijpend Lezen en Woordenschat

In tabel 4.12 hebben we de correlaties tussen de latente factoren weergegeven, zoals die met behulp van basismodel B zijn geschat. In bijlage 16 staan de ongestandaardiseerde parameterschattingen van model B voor alle negen datasets weergegeven.

De correlaties tussen EMT en AVI hebben we reeds besproken in paragraaf 4.2.3. Terugkomend op de relatie tussen de factoren Begrijpend Lezen en Woordenschat, kunnen we in tabel 4.12 zien dat deze zeer hoog is; veelal hoger dan de correlaties tussen EMT en AVI. Gemiddeld ligt de correlatie rond 0,90, met als uitschieters naar boven de eerste twee perioden van cohort 3 en de eerste periode van cohort 5.

Tabel 4.12: Correlaties tussen de factoren in basismodel B

	cohort 3			cohort 4			cohort 5		
	voorjaar groep 3	najaar groep 4	voorjaar groep 4	voorjaar groep 4	najaar groep 5	voorjaar groep 5	voorjaar groep 5	najaar groep 6	voorjaar groep 6
EMT - AVI	0,855	0,861	0,892	0,956	0,875	0,906	0,899	0,807	0,830
EMT - BL	0,630	0,716	0,605	0,611	0,585	0,512	0,438	0,528	0,455
EMT - WS	0,574	0,717	0,562	0,521	0,574	0,508	0,448	0,519	0,479
AVI - BL	0,560	0,687	0,636	0,667	0,596	0,591	0,551	0,535	0,514
AVI - WS	0,552	0,692	0,603	0,562	0,580	0,596	0,529	0,470	0,462
BL - WS	0,958	0,981	0,927	0,905	0,911	0,894	0,986	0,891	0,890

De zesde onderzoeksvraag betrof de relatie tussen AVI enerzijds en Begrijpend Lezen en Woordenschat anderzijds. Wat onmiddellijk duidelijk wordt uit de tabel is dat de AVI-toetskaarten duidelijk iets anders meten dan Begrijpend Lezen of Woordenschat: de correlaties met deze factoren zijn beduidend lager dan de correlaties met de factor EMT. De correlaties tussen de factoren AVI en Begrijpend Lezen zijn over het algemeen van dezelfde orde van grootte als die tussen AVI en Woordenschat.

Wanneer we de hoogte van de correlaties tussen AVI enerzijds en Begrijpend Lezen en Woordenschat anderzijds bekijken, dan lijkt er sprake te zijn van de volgende trend. De correlaties zijn het hoogst in groep 4 en nemen af in groep 5 en groep 6. Een vergelijkbare trend zien we voor de correlaties tussen EMT en beide factoren. Deze resultaten komen overeen met de bevinding uit andere onderzoeken dat de correlaties tussen technisch lezen en begrijpend lezen het hoogst zijn in de lagere leerjaren van het basisonderwijs (Aarnoutse, Mommers, Smits & Van Leeuwe, 1986; Boland, 1991; Daneman, 1991; zie ook paragraaf 1.4.1).

Wanneer we correlaties van de AVI-factor met Begrijpend Lezen en Woordenschat vergelijken met die van de EMT-factor met deze factoren, dan valt het volgende op. In het voorjaar van groep 3 en het najaar van groep 4 zijn de correlaties van EMT met beide factoren iets hoger dan die van AVI met deze factoren. In de latere perioden is dit voor de correlaties met Begrijpend Lezen omgekeerd: de correlaties van AVI met deze factoren is iets hoger. De correlaties met Woordenschat met de AVI-factor zijn iets hoger dan die met de EMT-factor in het voorjaar van groep 4 en het voor- en najaar van groep 5. In groep 6 zijn de correlaties met de EMT-factor weer iets hoger. Van een duidelijke trend lijkt geen sprake te zijn.

4.4 De relatie met begrijpend lezen en woordenschat: longitudinaal

In de vorige paragraaf zijn we per cohort per periode nagegaan wat de relatie is van de AVI-toetskaarten met toetsen voor begrijpend lezen en woordenschat. In deze paragraaf gaan we na hoe deze relatie longitudinaal is. We hebben de analyses apart uitgevoerd voor Begrijpend Lezen en voor Woordenschat. De hoge correlaties tussen deze beide latente variabelen (zie tabel 4.12) zouden problemen kunnen opleveren bij het toetsen van modellen waarin beide factoren zijn opgenomen.

In onderstaande modellen spreken we telkens over de relatie tussen AVI en Begrijpend Lezen, maar voor de relatie met Woordenschat hebben we dezelfde uitgangsmoellen gehanteerd.

4.4.1 De basismodellen

In de theorie omtrent leesvaardigheid wordt veelal verondersteld dat de technische leesvaardigheid een voorwaarde is voor het begrijpen van teksten (Perfetti, 1985; Stanovich, 1986; zie ook paragraaf 1.4.1). Longitudinaal onderzoek bevestigt dit beeld: kinderen die zwak zijn in technisch lezen zijn dat ook bij begrijpend lezen. Vooral in het begin van het leesonderwijs beïnvloeden de prestaties in het technisch lezen het leesbegrip. Vanuit de gedachte dat vorderingen in de technische leesvaardigheid een positieve invloed hebben op de vaardigheid in het begrijpend lezen, hebben we het basismodel C geformuleerd (zie figuur 4.4). In dit zogenaamde *veranderingsmodel* nemen we aan dat een verandering in de prestatie op de AVI-toets (=technisch lezen) tussen periode X en X+1 de prestatie op begrijpend lezen beïnvloedt.

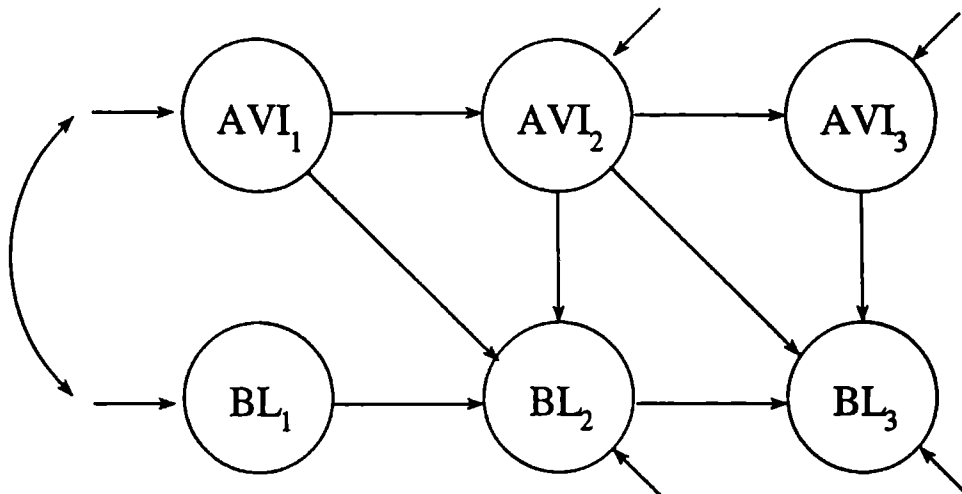
Zoals in de figuur te zien is, laten we in de eerste periode de correlatie tussen de factoren AVI en Begrijpend Lezen vrij. Voor periode 2 en 3 veronderstellen we ten aanzien van de relatie tussen de latente variabelen, dat:

- 1 de factor AVI in periode X+1 beïnvloed wordt door de factor AVI in periode X;
- 2 de factor Begrijpend Lezen in periode X+1 beïnvloed wordt door de factor Begrijpend Lezen in periode X, en
- 3 de factor Begrijpend lezen in periode X+1 beïnvloed wordt door de factor AVI in de perioden X en X+1.

We hebben deze laatste veronderstelling opgenomen in model C door de parameters behorende bij de invloed van AVI in de perioden X en X+1 naar Begrijpend Lezen in periode X+1 vrij te schatten.

De meeteenheid van de latente variabelen hebben we vastgelegd door middel van één van de geobserveerde variabelen. De latente variabele Begrijpend Lezen wordt in alle drie perioden gemeten met behulp van twee geobserveerde variabelen (zie ook paragraaf 4.3). In bijlage 11 staat per cohort per periode weergegeven welke toetsen dat zijn. Wanneer in twee perioden dezelfde toets voor begrijpend lezen is gebruikt veronderstellen we gecorreleerde errorvarianties tussen deze twee toetsen.

Voor de latente variabele AVI beschikken we in de eerste en de derde periode over de toetsgegevens van beide versies en in de tweede periode alleen over versie A van de AVI-toetskaarten. Op basis van de resultaten uit paragraaf 4.2 veronderstellen we dat de twee versies van de AVI-toets in de eerste en de derde periode parallel zijn (gelijke fac-



Figuur 4.4: Veranderingsmodel C voor de longitudinale relatie tussen AVI en Begrijpend Lezen (Woordenschat)

torladingen en errorvarianties). In tegenstelling tot de modellen in de vorige paragrafen kunnen we de errorvariantie van versie A van de AVI-toetskaarten in de tweede periode vrijlaten.

Aan model C hebben we vervolgens de beperking opgelegd dat de factor Begrijpend Lezen in periode $X+1$ beïnvloed wordt door *het verschil in AVI-factor tussen de perioden $X+1$ en X* . Met andere woorden: de *winst* in technisch lezen, gemeten met behulp van de AVI-toets beïnvloedt, onafhankelijk van het niveau, de prestaties in begrijpend lezen. In formule 4.2 hebben we dit weergegeven.

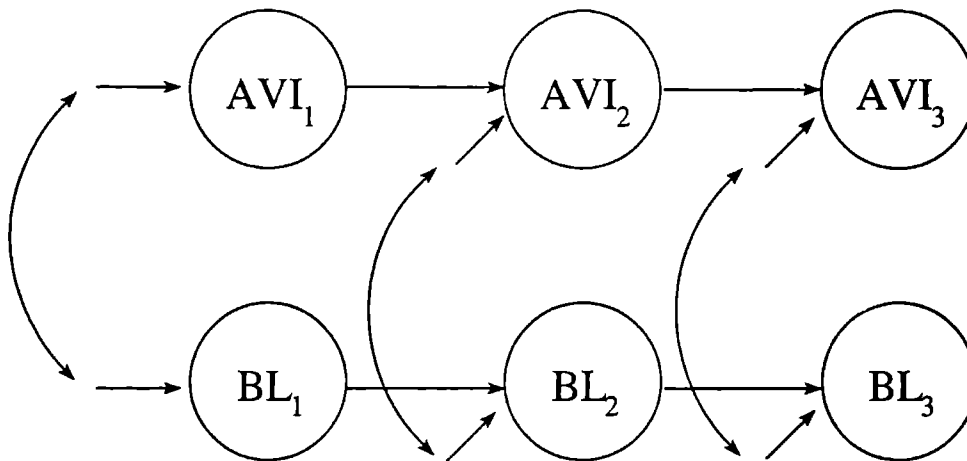
$$BL_{X+1} = \beta (AVI_{X+1} - AVI_X) = \beta AVI_{X+1} - \beta AVI_X \quad (4.2)$$

Uit deze formule blijkt dat om deze beperking aan model C toe te voegen, de parameter behorend bij de invloed van de factor AVI in periode X op Begrijpend Lezen in periode $X+1$ tegengesteld dient te zijn aan de parameter behorend bij de invloed van AVI-factor in periode $X+1$ op Begrijpend Lezen in periode $X+1$. Dit resulteert in model C1.

Het tweede longitudinale basismodel dat we toetsen is het *stabiliteitsmodel* (model D). In figuur 4.5 hebben we dit model weergegeven. In model D veronderstellen we, net als in model C, dat elke latente factor in periode $X+1$ beïnvloed wordt door dezelfde factor in periode X . Naast de twee eerste assumpties van model C veronderstellen we in model D dat:

3 de factoren AVI en Begrijpend Lezen onderling correleren in periode X .

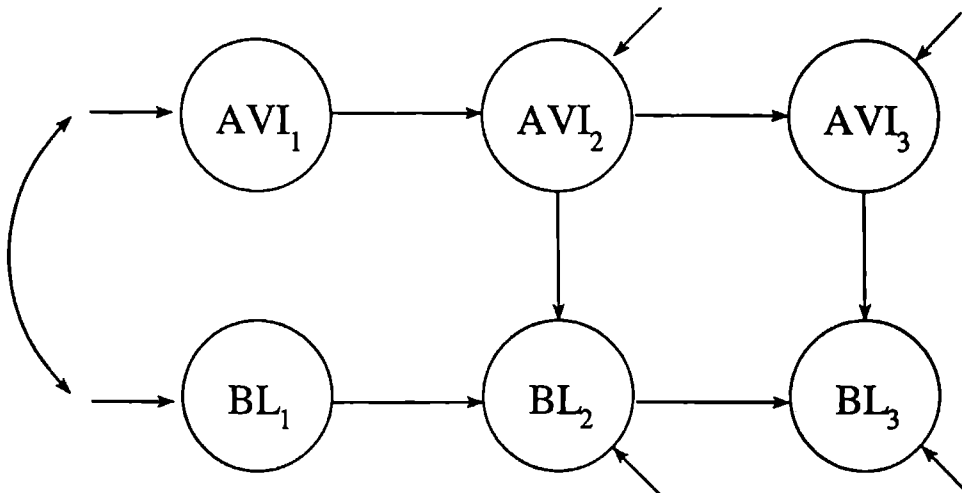
De assumpties ten aanzien van de geobserveerde variabelen en de meeteenheid van de latente variabelen zijn gelijk aan die in het veranderingsmodel C.



Figuur 4.5: Stabiliteitsmodel D voor de longitudinale relatie tussen AVI en Begrijpend Lezen (Woordenschat)

Het verschil met model C is dat we in model D geen causale relatie veronderstellen tussen AVI en Begrijpend Lezen. Wel wordt in het model verondersteld dat AVI en Begrijpend Lezen met elkaar samenhangen, waarbij geldt dat de correlatie tussen AVI en Begrijpend Lezen in periode $X+1$ niet volledig verklaard kan worden door de samenhang tussen beide factoren in periode X . We zullen tevens een variant van model D toetsen, waarin we volledige stabiliteit veronderstellen. In dit model (D1) veronderstellen we alleen een correlatie tussen AVI en Begrijpend Lezen in de eerste periode. De correlaties tussen de residuen van de factoren in de andere twee perioden zijn op gefixeerd op 0. Daarmee veronderstellen we dat de samenhang tussen beide factoren in de perioden 2 en 3 volledig wordt bepaald door de correlatie in periode 1. De implicatie van dit model is dat technisch lezen en begrijpend lezen met elkaar samenhangen, maar dat die samenhang vervolgens niet verandert.

Het laatste basismodel dat we zullen toetsen is het model dat in eerder Nederlands onderzoek veelvuldig gebruikt is. Het model is ontwikkeld met de gegevens van het project 'Preventie van Leesmoeilijkheden' (Boland, 1991). In figuur 4.6 hebben we dit model weergegeven, dat we in het vervolg het *Preventie-model* (model E) noemen. Model E kan als een variant worden gezien van model C. Het verschil tussen de modellen is dat in model D verondersteld wordt dat D alleen binnen eenzelfde periode de AVI-factor invloed heeft op de factor Begrijpend Lezen. Het bereikte niveau in technisch lezen in periode $X+1$, gemeten met de AVI-toets, beïnvloedt de prestatie in begrijpend lezen in dezelfde periode. Er is geen sprake van dat de winst in technische leesvaardigheid tussen de perioden X en $X+1$ een effect heeft op de vaardigheid in begrijpend lezen in periode $X+1$.



Figuur 4.6: 'Preventie-model' E voor de longitudinale relatie tussen AVI en Begrijpend Lezen (Woordenschat)

De twee eerste assumpties van model C ten aanzien van het verband tussen de latente factoren gelden ook voor dit model. De derde veronderstelling is vervangen door de assumptie dat:

- 3 de factor Begrijpend Lezen in periode X beïnvloed wordt door de factor *AVI in periode X*.

De assumpties ten aanzien van de geobserveerde variabelen en de meeteenheid van de latente variabelen zijn gelijk aan die in de andere twee modellen.

4.4.2 De longitudinale relatie tussen AVI en Begrijpend Lezen

Bovenstaande longitudinale modellen voor de relatie tussen AVI en Begrijpend Lezen zijn afzonderlijk getoetst op de data van elk cohort. Afhankelijk van de uitkomsten van de modeltoetsing zijn de betreffende modellen nog enigszins aangepast.

In tabel 4.13 hebben we de resultaten van de modeltoetsing weergegeven voor de relatie tussen AVI en Begrijpend Lezen. Per cohort presenteren we eerst de hierboven besproken modellen en vervolgens de aanpassingen van die modellen.

Tabel 4.13: Toetsing longitudinale modellen voor de relatie tussen AVI en Begrijpend Lezen

	χ^2	df	p	RMSEA	$p_{\text{close fit}}$	GFI	CFI	BIC
COHORT 3 (N=859)								
C Veranderingsmodel	129,778	37	0,000	0,054	0,241	0,973	0,989	-208,908
C1=C + (AVI _t ->BL _{t+1})=- -(AVI _{t+1} -> BL _{t+1})	147,827	39	0,000	0,057	0,111	0,969	0,987	-209,166
D Stabiliteitsmodel	158,669	39	0,000	0,060	0,045	0,967	0,985	-198,324
E 'Preventie'-model	188,115	39	0,000	0,067	0,002	0,962	0,982	-168,878
C2=C1 + BL _t ->AVI _{t+1}	71,115	37	0,001	0,033	0,995	0,985	0,996	-267,571
C2a=als C2, maar alleen BL _t ->AVI _t	76,777	38	0,000	0,035	0,991	0,984	0,995	-270,062
C3=C2a + tussen periode 2 en 3 alleen stabiliteitscoëfficiënten	90,093	39	0,000	0,039	0,995	0,981	0,994	-266,900
D1 alleen correlatie tussen AVI en BL in periode 1	269,565	41	0,000	0,081	0,000	0,950	0,972	-105,735
COHORT 4 (N=796)								
C Veranderingsmodel	76,949	37	0,000	0,037	0,970	0,983	0,995	-258,918
C1=C + (AVI _t ->BL _{t+1})=- -(AVI _{t+1} -> BL _{t+1})	77,062	39	0,000	0,035	0,986	0,983	0,995	-276,960
D Stabiliteitsmodel	80,436	39	0,000	0,037	0,976	0,982	0,995	-273,586
E 'Preventie'-model	97,973	39	0,000	0,044	0,826	0,978	0,993	-256,049
C2=C1 + BL _t ->AVI _{t+1}	67,068	37	0,002	0,032	0,994	0,985	0,996	-268,799
C2a=als C2, maar alleen BL _t ->AVI _t	70,582	38	0,001	0,033	0,993	0,984	0,996	-274,363
C3=C1 + tussen periode 2 en 3 alleen stabiliteitscoëfficiënten	86,938	40	0,000	0,038	0,958	0,980	0,994	-276,162
D1 alleen correlatie tussen AVI en BL in periode 1	98,721	41	0,000	0,042	0,885	0,978	0,993	-273,456
COHORT 5 (N=739)								
C Veranderingsmodel	59,014	37	0,012	0,028	0,998	0,986	0,997	-274,104
C1=C + (AVI _t ->BL _{t+1})=- -(AVI _{t+1} -> BL _{t+1})	60,893	39	0,014	0,028	0,999	0,985	0,997	-290,232
D Stabiliteitsmodel	60,318	39	0,016	0,027	0,999	0,985	0,997	-290,807
E 'Preventie'-model	60,011	39	0,017	0,027	0,999	0,985	0,997	-291,114
D1 alleen correlatie tussen AVI en BL in periode 1	62,092	41	0,018	0,026	0,999	0,985	0,997	-307,039

De resultaten van de toetsing van de modellen C tot en met E voor *cohort 3* wijzen op het volgende. Volgens de χ^2 -toets past geen van de modellen bij de data. De toetsing van de close fit van de RMSEA geeft een iets ander beeld. Volgens deze toetsing zouden

de twee veranderingsmodellen wel bij de data passen (model C en model C1), maar het stabiliteitsmodel D en het 'Preventie'-model E niet. De hoogte van de GFI en de CFI van deze laatste twee modellen wijst er overigens wel op dat deze modellen goed bij de data zouden passen.

Op basis van de BIC zou model C1 de voorkeur hebben boven de andere modellen. De toetsing van het verschil in χ^2 tussen C en C1 daarentegen geeft aan dat model C beter bij de data past ($\chi^2_{C1-C}=18,0$, $df=2$, $p<0,01$). We geven de voorkeur aan het meest eenvoudige model C1, waarin verondersteld wordt dat een toename in de technische leesvaardigheid (gemeten met de AVI-toetskaarten) de prestatie op begrijpend lezen beïnvloedt.

Op basis van de modificatie-indices hebben we model C1 nog verder aangepast. In model C2 hebben we de parameters vrijgelaten betreffende de invloed van Begrijpend Lezen in periode 1 op AVI-factor in periode 2 en van Begrijpend Lezen in periode 2 op AVI-factor in periode 3. We veronderstellen als het ware dat het niveau van begrijpend lezen in een periode het niveau van technisch lezen in een volgende periode bepaalt. In model C2a hebben we alleen de parameter voor de invloed van Begrijpend Lezen in periode 1 op de factor AVI in periode 2 vrijgelaten.

Toetsing van het verschil in χ^2 tussen deze modellen en model C1 geeft aan dat het vrijlaten van een of beide parameters tot een duidelijke verbetering van de modelpassing leidt ($\chi^2_{C1-C2}=76,7$, $df=2$, $p<0,01$; $\chi^2_{C1-C2a}=71,1$, $df=1$, $p<0,01$). De BIC van deze twee modellen is duidelijk lager dan de BIC van model C1. Op basis van de BIC zou model C2a de voorkeur hebben boven model C2, op basis van het verschil in χ^2 zou model C2 de voorkeur hebben boven model C2a. We geven de voorkeur aan het meest eenvoudige model C2a. In dit model wordt verondersteld dat de prestatie op Begrijpend Lezen in periode $X+1$ wordt beïnvloed door de prestatie op Begrijpend Lezen in periode X én door de toename in de prestatie op Technisch Lezen (gemeten met de AVI-toets) tussen de perioden X en $X+1$. Daarnaast geldt voor het begin van het leesonderwijs dat de prestatie op Begrijpend Lezen in periode 1 de prestatie op Technisch Lezen in periode 2 beïnvloedt. In tabel 4.13 staan bij cohort 3 nog twee andere modellen vermeld. De resultaten hiervan zullen we in een later stadium bespreken.

De resultaten voor *cohort 4* in tabel 4.13 laten het volgende beeld zien. Geen van de modellen C tot en met E passen volgens de χ^2 -toets bij de data; de overige fitmaten daarentegen geven aan dat alle modellen bij de data passen. Wanneer we de BIC van de modellen vergelijken, dan zou de voorkeur uitgaan naar model C1, het veranderingsmodel. Model C1 hebben we dan ook als basis gekozen voor verdere aanpassingen.

Analoog aan de modelontwikkeling voor cohort 3 hebben we eerst parameters toegevoegd die de invloed weergegeven van Begrijpend Lezen in periode X naar de AVI-factor in periode $X+1$ (modellen C2 en C2a). Zowel uit de toetsing van verschillen in χ^2 ($\chi^2_{C1-C2}=10,0$, $df=2$, $p<0,01$; $\chi^2_{C1-C2a}=6,5$, $df=1$, $p<0,02$) als uit de vergelijking van de BIC's blijkt dat het meest restrictieve model C1 de voorkeur geniet boven de modellen C2 en C2a.

Model C1 hebben we vervolgens nog verder vereenvoudigd door te veronderstellen dat tussen periode 2 en 3 alleen nog sprake is van stabiliteitscoëfficiënten voor de AVI-fac-

tor en de factor Begrijpend Lezen. In het resulterende model C3 veronderstellen we dat Begrijpend lezen in periode 3 alleen beïnvloed wordt door Begrijpend lezen in periode 2. Toetsing van het verschil in χ^2 tussen model C1 en C3 geeft aan dat de voorkeur uitgaat naar model C1 ($\chi^2_{C3-C1}=9,9$, $df=1$, $p<0,01$). De BIC van beide modellen is even groot, hetgeen betekent dat beide modellen gelijkwaardig zijn. In paragraaf 4.1.2 hebben we aangegeven dat in zo'n geval de t-waarde van de parameter die in het ene model wel en in het andere niet is opgenomen, doorslaggevend is. De t-waarde behorende bij de invloed van de toename in de factor AVI tussen periode 2 en 3 op Begrijpend Lezen in periode 3 bedraagt 2,78. Dit betekent dat sprake is van een zwakke aanwijzing voor deze invloed. Voor een sterke aanwijzing dient de t-waarde van deze parameter groter of gelijk aan 3,42 te zijn ($t=\sqrt{[\ln(796)+5]}$; zie paragraaf 4.1.2). Gezien dit resultaat geven we de voorkeur aan het meest eenvoudige model C3: tussen periode 1 en 2 heeft een toename in AVI-niveau nog een effect op het niveau van begrijpend lezen, tussen periode 2 en 3 zijn de verschillen gestabiliseerd.

Tot slot hebben we model D1 getoetst, waarin voor de latente variabelen alleen sprake is van volledige stabiliteit tussen de perioden 1 en 2 en de perioden 2 en 3. De AVI-factor en Begrijpend Lezen zijn wel gecorreleerd in de eerste periode. In dit model wordt verondersteld dat de samenhang tussen Technisch Lezen en Begrijpend Lezen al bestaat vanaf het begin en dat daarin vervolgens niets meer verandert. Toetsing van het verschil in χ^2 tussen de modellen C3 en D1 geeft aan dat model C3 de voorkeur heeft. De BIC van het meer restrictieve model D1 is hoger dan de BIC van model C3, maar het verschil bedraagt minder dan 6 punten zodat we op basis daarvan geen keuze voor een van de twee modellen kunnen maken: beide modellen zijn gelijkwaardig. De t-waarde behorende bij de invloed van een toename in technische leesvaardigheid tussen periode 1 en 2 op de factor Begrijpend Lezen in periode 2 bedraagt 3,32, terwijl voor een sterke aanwijzing voor deze parameter de t-waarde groter of gelijk aan 3,42 dient te zijn. Dit betekent dat er een zwakke aanwijzing is voor de parameter.

Alle resultaten overziend betekent dit dat we aan model D1 de voorkeur dienen te geven. Toch willen we model C3 niet meteen verwerpen ten voordele van model D1, mede omdat de t-waarde van de parameter in model C3 maar net niet voldoet voor een sterke aanwijzing. We willen eerst de resultaten voor cohort 5 en voor de relatie tussen AVI en Woordenschat bekijken, alvorens een keuze te maken.

Gezien de modeluitkomsten voor cohort 4 hebben we voor cohort 3 ook nog enkele modellen getoetst, waarvan we er twee weergeven in tabel 4.13. Het laatste model D1, met alleen maar stabiliteitscoëfficiënten tussen de AVI-factoren respectievelijk de factoren van Begrijpend Lezen in de tijd, past duidelijk niet bij de data, noch volgens de χ^2 -toets, noch volgens de close fit van de RMSEA. De BIC is tevens duidelijk hoger dan de BIC van de andere modellen. Model C3, waarbij tussen periode 2 en 3 uitsluitend sprake is van stabiliteitscoëfficiënten voor beide latente variabelen, past volgens de RMSEA bij de data. De toetsing van het verschil in χ^2 tussen C3 en C2a valt uit in het voordeel van model C2a. De BIC van het meer restrictieve model C3 is hoger dan de BIC van C2a, maar het verschil bedraagt minder dan 6 punten. De t-waarde van de invloed van de toename in technische leesvaardigheid tussen de tweede en derde periode op de vaardigheid in

begrijpend lezen in de derde periode bedraagt 3,64. Dit kan gezien worden als een sterke aanwijzing voor de betreffende parameter, zodat we voor cohort 3 de voorkeur blijven geven aan model C2a.

Tot slot staan in tabel 4.13 de resultaten voor *cohort 5* weergegeven. De toetsingsresultaten voor de modellen C tot en met E verschillen niet of nauwelijks van elkaar. Geen van de modellen past volgens de χ^2 -toets (waarvan de waarden vrijwel gelijk zijn) en alle modellen passen volgens de andere fit-maten. In het laatste model D1 veronderstellen we een correlatie tussen de factoren AVI en Begrijpend Lezen in de eerste periode en vervolgens alleen autocorrelaties voor beide factoren. Dit model past eveneens goed bij de data volgens alle fitmaten behalve de χ^2 . De BIC van dit meest eenvoudige model is duidelijk het laagst, zodat we aan dit model de voorkeur geven.

In deze paragraaf hebben we modellen getoetst voor de longitudinale relatie tussen technisch lezen gemeten met de AVI-toetskaarten en begrijpend lezen. Alvorens tot de uiteindelijke interpretatie van de resultaten te komen willen we eerst dezelfde modellen toetsen voor de relatie tussen technisch lezen en woordenschat.

4.4.3 De longitudinale relatie tussen AVI en Woordenschat

De longitudinale modellen zoals gepresenteerd in de figuren 4.4 tot en met 4.6 hebben we eveneens getoetst voor de relatie tussen AVI en Woordenschat. Ook nu zijn de modellen afzonderlijk per cohort toegepast. In tabel 4.14 staan de resultaten van de modeltoetsingen weergegeven.

Wanneer we de resultaten van de modeltoetsing voor *cohort 3* bekijken, zien we het volgende. De modellen C tot en met E passen volgens de χ^2 -toets niet bij de data, maar volgens de overige fit-maten wel. Een opvallend verschil met de resultaten voor de relatie tussen AVI en Begrijpend Lezen is dat model C1 veel slechter bij de data past dan model C. De χ^2 en de BIC van model C1 zijn veel hoger dan die van model C. Model C past ook duidelijk veel beter bij de data van cohort 3 dan de modellen D en E. Het gegeven dat model C beter bij de data past dan model C1 geeft aan dat voor woordenschat geldt dat de invloed van de winst in technische leesvaardigheid tussen de perioden X en X+1 op Woordenschat in periode X+1 niet onafhankelijk is van het niveau van technisch lezen. Dit resultaat plaatst ons voor een probleem, aangezien we zouden verwachten dat de modellen voor begrijpend lezen en woordenschat vergelijkbaar zouden zijn. Op dit moment willen we nog geen keuze voor een van de twee modellen maken. Voor de verdere modelaanpassingen hebben we daarom beide modellen als uitgangspunt genomen.

Tabel 4.14: Toetsing longitudinale modellen voor de relatie tussen AVI en Woordenschat

	χ^2	df	p	RMSEA	$p_{close\ fit}$	GFI	CFI	BIC
COHORT 3 (N=859)								
C Veranderingsmodel	86,056	37	0,000	0,039	0,946	0,982	0,994	-252,630
C1=C + (AVI _t ->WS _{t+1})=- -(AVI _{t+1} -> WS _{t+1})	140,644	39	0,000	0,055	0,185	0,970	0,988	-216,349
D Stabiliteitsmodel	157,102	39	0,000	0,059	0,052	0,967	0,986	-199,891
E 'Preventie'-model	124,272	39	0,000	0,051	0,450	0,975	0,990	-232,721
C2=C1 + WS _t ->AVI _{t+1}	110,397	37	0,000	0,048	0,602	0,977	0,991	-228,289
C2a=als C2, maar alleen WS ₁ ->AVI ₂	116,109	38	0,000	0,049	0,549	0,976	0,991	-231,730
C3=C2a + tussen perioden 2-3 alleen stabiliteitscoefficienten	118,848	39	0,000	0,049	0,557	0,975	0,990	-238,145
D1 alleen correlatie tussen AVI en WS in periode 1	263,398	41	0,000	0,080	0,000	0,951	0,973	-111,902
<hr/>								
C4=C + WS ₁ ->AVI ₂	52,837	36	0,035	0,023	1,000	0,989	0,998	-276,695
C5=C4 + tussen perioden 2-3 alleen stabiliteitscoefficienten	86,273	38	0,000	0,039	0,961	0,982	0,994	-261,566
C5a=C5 + AVI ₂ ->WS ₃	58,395	37	0,014	0,026	1,000	0,988	0,997	-280,291
C5b=C5 + AVI ₃ ->WS ₃	69,988	37	0,001	0,032	0,996	0,985	0,996	-268,698
COHORT 4 (N=796)								
C Veranderingsmodel	115,822	37	0,000	0,052	0,374	0,973	0,991	-220,045
C1=C + (AVI _t ->WS _{t+1})=- -(AVI _{t+1} -> WS _{t+1})	123,918	39	0,000	0,052	0,339	0,972	0,991	-230,104
D Stabiliteitsmodel	122,234	39	0,000	0,052	0,369	0,972	0,991	-231,788
E 'Preventie'-model	129,630	39	0,000	0,054	0,244	0,970	0,990	-224,392
C3=C1 + tussen perioden 2-3 alleen stabiliteitscoefficienten	130,070	40	0,000	0,053	0,287	0,970	0,990	-233,030
D1 alleen correlatie tussen AVI en WS in periode 1	138,843	41	0,000	0,055	0,204	0,969	0,989	-233,334
<hr/>								
COHORT 5 (N=739)								
C Veranderingsmodel	80,160	38	0,000	0,039	0,940	0,980	0,995	-261,961
C1=C + (AVI _t ->BL _{t+1})=- -(AVI _{t+1} -> BL _{t+1})	83,083	40	0,000	0,038	0,953	0,980	0,995	-277,045
D Stabiliteitsmodel	83,502	40	0,000	0,038	0,951	0,980	0,995	-276,626
E 'Preventie'-model	81,789	40	0,000	0,038	0,961	0,980	0,995	-278,339
D1 alleen correlatie tussen AVI en WS in periode 1	84,445	42	0,000	0,037	0,971	0,980	0,995	-293,689

Model C1 hebben we op dezelfde manier aangepast als voor de relatie tussen AVI en Begrijpend Lezen. In model C2 hebben we de parameters vrijgelaten voor de invloed van

Woordenschat in periode 1 respectievelijk 2 op de AVI-factor in periode 2 respectievelijk 3. In model C2a hebben we alleen de parameter van Woordenschat in periode 1 op de AVI-factor in periode 2 vrijgelaten. Wanneer we model C1 vergelijken met de modellen C2 en C2a dan geven de verschillen in χ^2 aan dat beide laatste modellen duidelijk beter bij de data passen ($\chi^2_{C1-C2}=30,2$, $df=2$, $p<0,01$; $\chi^2_{C1-C2a}=24,5$, $df=1$, $p<0,01$). De BIC's van beide modellen zijn tevens duidelijk lager dan die van model C1. Wanneer we toetsingsresultaten van de modellen C2 en C2a met elkaar vergelijken, dan zou op basis van het verschil in χ^2 model C2 de voorkeur hebben boven C2a. Op basis van de BIC daarentegen zou model C2a de voorkeur genieten. We geven de voorkeur aan het meest eenvoudige model C2a, net als bij de modellen voor de relatie tussen AVI en Begrijpend lezen: de woordenschat in periode 1 beïnvloedt de prestaties in technisch lezen in periode 2.

Tot slot hebben we, analoog aan paragraaf 4.4.2, nog de modellen C3 (tussen perioden 2 en 3 volledige stabiliteit) en D1 (correlatie tussen AVI-factor en Woordenschat in periode 1 en stabiliteitscoëfficiënten voor beide factoren tussen de perioden) getoetst. Model D1 past duidelijk niet bij de data volgens de χ^2 -toets en de toetsing van de close fit van de RMSEA, hetgeen betekent dat de samenhang tussen AVI en Woordenschat in periode 1 niet voldoende is om de samenhang tussen beide factoren in de overige twee perioden te verklaren. De BIC van dit model is eveneens veel hoger dan dat bij de andere modellen. Model C3 heeft zowel volgens de toetsing van het verschil in χ^2 ($\chi^2_{C3-C2a}=2,7$, $df=1$, $p>0,05$) als volgens de BIC de voorkeur boven model C2a. Dit betekent dat de winst in de technische leesvaardigheid tussen de perioden 2 en 3 geen effect heeft op de woordenschat in periode 3. Deze uitkomst voor de relatie tussen AVI en Woordenschat verschilt met die voor de relatie tussen AVI en Begrijpend Lezen, waarbij model C2a de voorkeur had.

Gezien de uitkomsten van de vergelijking van de modellen C en C1 hebben we *model C* op een aantal manieren aangepast. In model C4 hebben we de invloed toegevoegd van de factor Woordenschat in periode 1 op de AVI-factor in periode 2 (vergelijkbaar met model C2a). Dit model past duidelijk beter bij de data dan model C. Vervolgens hebben we de restrictie opgelegd dat tussen periode 2 en 3 alleen sprake is van stabiliteit voor beide factoren (vergelijkbaar met model C3). Het resulterende model C5 past duidelijk minder goed bij de data dan model C4. Dit betekent dat alleen stabiliteitscoëfficiënten tussen de perioden 2 en 3 niet voldoende zijn: de samenhang tussen woordenschat en technische leesvaardigheid in de voorgaande perioden is niet voldoende om de samenhang in periode 2 te verklaren. Deze uitkomst is tegenstrijdig met de uitkomst van de vergelijking van de modellen C2a en C3.

In de modellen C5a en C5b hebben we de restricties in model C5 stapsgewijs ingevoerd. In model 5a veronderstellen we, vergeleken met model C4, alleen een invloed van de AVI-factor in periode 2 op de factor Woordenschat in periode 3. In model 5b veronderstellen we alleen een invloed van de AVI-factor in periode 2 op de factor Woordenschat in dezelfde periode. Model 5b past slechter bij de data dan model 4, zowel gezien het verschil in χ^2 ($\chi^2_{C5b-C4}=17,2$, $df=1$, $p<0,01$) als gezien het verschil in BIC. Model 5a geniet de voorkeur boven model C4 gezien de lagere BIC, maar niet volgens het verschil in χ^2 ($\chi^2_{C5a-C4}=5,6$, $df=1$, $p<0,02$).

Alle modeltoetsingen overziend ten aanzien van de relatie tussen AVI en Woordenschat in cohort 3, kunnen we concluderen dat de resultaten niet geheel consistent zijn. Uitgaand van model C1 zouden we moeten concluderen dat een toename in Technische Leesvaardigheid tussen de tweede en derde periode geen invloed meer heeft op Woordenschat in de derde periode. Vertrekkend vanuit model C, dat beter bij de data paste dan model C1, zouden we moeten concluderen dat er nog wel degelijk een invloed is van Technisch Lezen in periode 2 op Begrijpend Lezen in periode 3. In tabel 4.15 hebben we de ongestandaardiseerde parameterschattingen weergegeven voor de relaties tussen de latente variabelen geschat met de modellen C2a, C3, C4 en C5a.

Tabel 4.15: Ongestandaardiseerde parameterschatting (tussen haakjes de bijbehorende standaardfouten) van de structurele relaties in vier longitudinale modellen voor de relatie tussen AVI en Woordenschat

	Model C2a	Model C3	Model C4	Model C5a
AVI ₁ -> AVI ₂	1,014 (0,044)	1,009 (0,044)	0,996 (0,040)	0,996 (0,040)
AVI ₁ -> WS ₂	-0,821 ^a (0,102)	-0,851 ^c (0,104)	-0,327 (0,114)	-0,366 (0,115)
WS ₁ -> AVI ₂	0,068 (0,014)	0,068 (0,013)	0,070 (0,012)	0,070 (0,012)
WS ₁ -> WS ₂	0,559 (0,031)	0,557 (0,031)	0,427 (0,030)	0,426 (0,031)
AVI ₂ -> WS ₂	0,821 ^a (0,102)	0,851 ^c (0,104)	0,912 (0,095)	0,940 (0,096)
AVI ₂ -> AVI ₃	0,976 (0,024)	0,986 (0,024)	0,993 (0,024)	0,996 (0,024)
AVI ₂ -> WS ₃	-0,442 ^b (0,252)	-	-1,303 (0,319)	-0,645 (0,131)
WS ₂ -> WS ₃	1,550 (0,060)	1,551 (0,060)	1,859 (0,094)	1,844 (0,093)
AVI ₃ -> WS ₃	0,442 ^b (0,252)	-	0,617 (0,267)	-

N.B.: de letters a t/m c geven aan welke parameters aan elkaar gelijkgesteld zijn

Met behulp van de parameterschattingen uit de tabel kunnen we de regressievergelijkingen uitschrijven voor de factor Woordenschat in de perioden 2 en 3. Voor de factor Woordenschat in periode 2 zien de vergelijkingen in de verschillende modellen er als volgt uit:

$$C2a: WS_2 = .559 WS_1 + .821 AVI_2 - .821 AVI_1$$

$$C3 : WS_2 = .557 WS_1 + .851 AVI_2 - .851 AVI_1$$

$$C4 : WS_2 = .427 WS_1 + .912 AVI_2 - .327 AVI_1$$

$$C5a: WS_2 = .426 WS_1 + .940 AVI_2 - .366 AVI_1$$

We kunnen het deel van deze vergelijkingen dat betrekking heeft op het effect van AVI op Woordenschat herschrijven, waarmee duidelijk wordt wat het verschil is tussen de modellen C2a en C3 enerzijds en de modellen C4 en C5a anderzijds. In de eerste twee modellen geldt dat $WS_2 = .8(AVI_2 - AVI_1)$. Voor de andere twee modellen geldt dat $WS_2 = .3(3AVI_2 - AVI_1)$, hetgeen betekent dat de meting van het AVI-niveau in de tweede periode (oktober groep 4) drie maal zo zwaar telt in de vergelijking. Er is niet alleen sprake van een invloed van de winst in technische leesvaardigheid, maar ook van het niveau van technisch lezen.

Voor de factor Woordenschat in periode 3 geldt volgens model C2a dat $WS_3 = .4(AVI_3 - AVI_2)$, terwijl model C4 aangeeft dat $WS_3 = .6(AVI_3 - 2AVI_2)$. Dit laatste resultaat is niet consistent met dat in de tweede periode. Daarvoor geldt dat AVI in de dezelfde periode als die waarin Woordenschat is gemeten zwaarder meetelt, terwijl voor Woordenschat in de derde periode juist de meting van AVI in de voorafgaande periode een groter gewicht krijgt.

De resultaten voor cohort 3 geven in ieder geval aan dat de invloed van de winst in technische leesvaardigheid op de woordenschat wellicht niet onafhankelijk is van het niveau van technisch lezen. Maar gezien de inconsistente resultaten is het niet mogelijk om op basis van de parameterschattingen een keuze te maken voor één van de modellen betreffende de relatie tussen AVI en Woordenschat. Bij de keuze van het uiteindelijke model voor cohort 3 hebben we ons gebaseerd op de resultaten voor de relatie tussen AVI en Begrijpend Lezen. Omdat deze resultaten wel eenduidig zijn en aangezien de factoren Begrijpend Lezen en Woordenschat in hoge mate gecorreleerd zijn, kiezen we ook voor de relatie van AVI met Woordenschat voor model C2a.

De resultaten van de modeltoetsing voor *cohort 4* ten aanzien van de relatie tussen de AVI-factor en Woordenschat leiden tot hetzelfde eindmodellen als voor de relatie tussen de AVI-factor en Begrijpend Lezen. Van de modellen C tot en met E heeft model C1 gezien de toetsingsresultaten de voorkeur. In dit model veronderstellen we dat een toename in de Technische Leesvaardigheid tussen twee perioden de Woordenschat in de tweede periode positief beïnvloedt. Vervolgens zijn de modellen C3 en D1 getoetst. In model C3 worden tussen de laatste twee perioden, in model D1 worden in alle perioden alleen stabiliteitscoëfficiënten tussen de latente variabelen worden verondersteld. De vergelijking van de modellen C1 en C3 geeft aan dat volgens het verschil in χ^2 model C1 de voorkeur geniet ($\chi^2_{C3-C1} = 6,2$, $df=1$, $p<0,02$), terwijl volgens het verschil in BIC model C3 de voorkeur krijgt. We geven de voorkeur aan het meest eenvoudige model C3, waarin verondersteld wordt dat een toename in technische leesvaardigheid tussen periode 2 en 3 geen invloed heeft op de woordenschat in periode 3.

Tot slot vergelijken we de modellen C3 en D1 met elkaar. Het verschil in χ^2 tussen de modellen C3 en D1 wijst erop dat model C3 de voorkeur van deze twee modellen geniet ($\chi^2_{D1-C3} = 8,8$, $df=1$, $p<0,01$). De BIC's van beide modellen zijn vrijwel identiek, zodat we de modellen als gelijkwaardig kunnen beschouwen. Inspectie van de t-waarde ($t=2,92$) behorend bij de invloed van het verschil in de AVI-factor tussen de eerste twee perioden op de factor Woordenschat in de tweede periode geeft aan dat sprake is van een zwakke aanwijzing voor deze relatie. Dit komt overeen met de resultaten voor relatie tussen AVI en Begrijpend Lezen.

De toetsing van de modellen C tot en met E voor *cohort 5* leverde een probleem bij de parameterschattingen op. De schatting van de residuele variantie van de factor Woordenschat in de tweede periode nam een negatieve waarde aan, waardoor de verklaarde variantie voor deze factor meer dan 100% bedroeg. Dit probleem hebben we opgelost door deze parameter gelijk te stellen aan 0,001, hetgeen nauwelijks een effect bleek te hebben op de overige parameterschattingen. In tabel 4.14 zijn de toetsingsresultaten van de modellen weergegeven waarbij deze restrictie is toegepast. Deze resultaten leiden tot hetzelfde uiteindelijke model voor de relatie tussen AVI en Woordenschat als voor de relatie tussen AVI en Begrijpend Lezen: model D1, waarin de factoren AVI en Woordenschat in de eerste periode met elkaar correleren. Tussen de perioden is uitsluitend sprake van stabiliteitscoëfficiënte. Technisch Lezen respectievelijk Woordenschat in periode X beïnvloeden Technisch Lezen respectievelijk Woordenschat in de periode X+1, maar een toename in de Technische Leesvaardigheid heeft geen invloed op de Woordenschat.

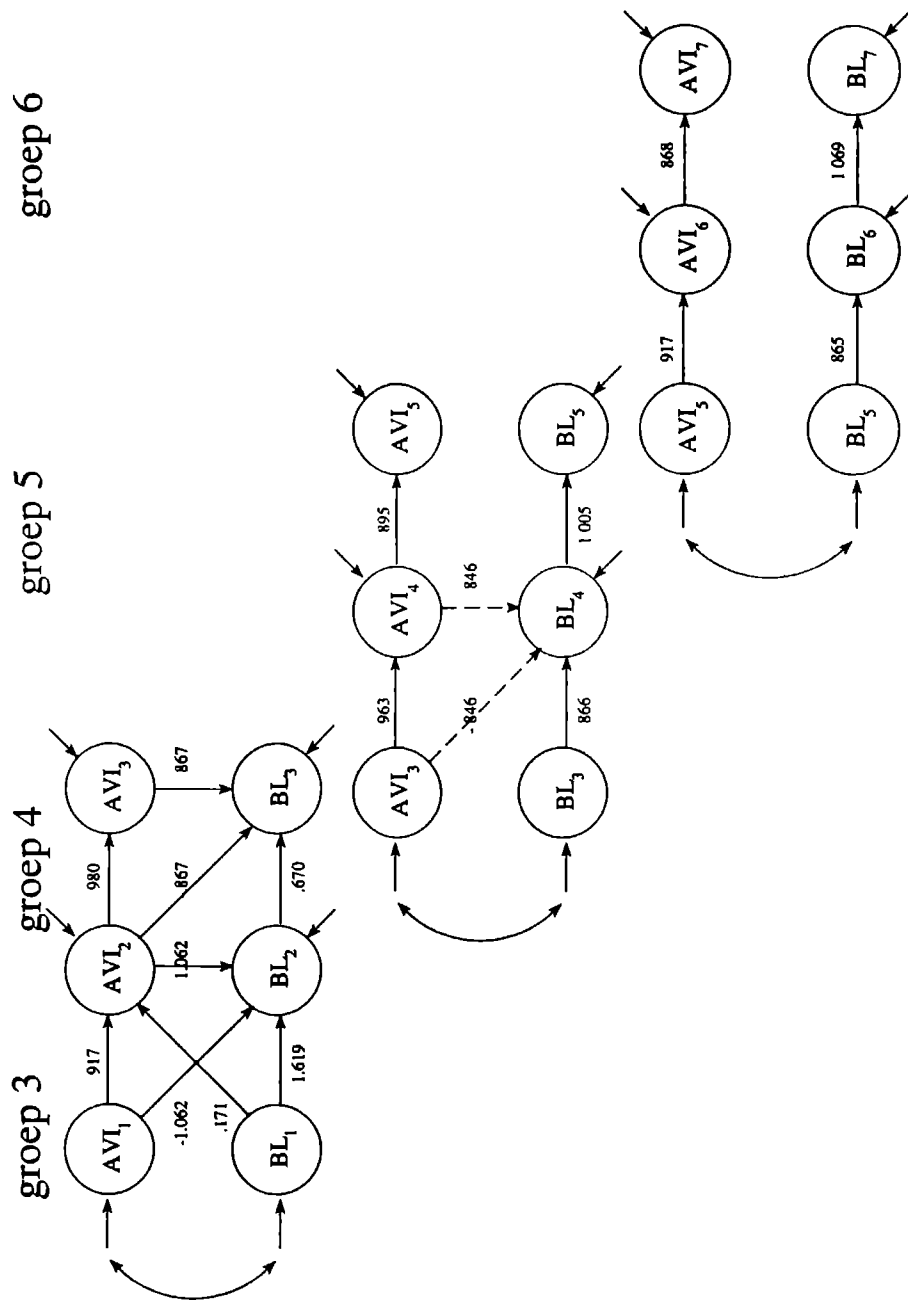
4.4.4 Interpretatie van de uiteindelijke modellen

In de paragrafen 4.4.2 en 4.4.3 hebben we de resultaten besproken betreffende de toetsingen van de longitudinale modellen. De uitkomsten voor de relatie tussen de factor AVI en Begrijpend Lezen zijn redelijk eenduidig en in overeenstemming met onze theoretische veronderstellingen: een toename in de technische leesvaardigheid beïnvloedt de vaardigheid in begrijpend lezen en deze invloed neemt af in de tijd. Daarentegen zijn de resultaten voor de relatie tussen de factoren AVI en Woordenschat minder eenduidig, althans voor *cohort 3*. Voor de overige cohorten geldt dat de uitkomsten overeenkomen met die voor de relatie met Begrijpend Lezen.

In figuur 4.7 hebben we gevisualiseerd hoe volgens ons de relatie tussen de factoren AVI en Begrijpend Lezen zich ontwikkelt in de basisschoolperiode van het voorjaar van groep 3 tot en met het voorjaar van groep 6. Voor de structurele relaties hebben we de ongestandaardiseerde parameterschattingen weergegeven in de figuur. In bijlage 17 staan alle ongestandaardiseerde parameterschattingen weergegeven van de hieronder genoemde modellen voor de relatie tussen de factor AVI en Begrijpend Lezen.

Het eerste deel van figuur 4.7 betreft model C2a van *cohort 3*. In dit model veronderstellen we dat een verbetering van de technische leesvaardigheid (gemeten met de AVI-toets) een positieve invloed heeft op de vaardigheid in begrijpend lezen. Zo beïnvloedt het verschil in technische leesvaardigheid tussen het voorjaar van groep 3 en het najaar van groep 4 (tijdstip 1 en 2 in de figuur) de vaardigheid in het begrijpend lezen in het najaar van groep 4. De winst in de technische leesvaardigheid tussen twee perioden heeft een positieve invloed op de begrijpende leesvaardigheid, hetgeen overeenkomt met de theoretische veronderstelling dat een goede technische leesvaardigheid een positieve invloed heeft op het goed kunnen begrijpen van teksten (zie paragraaf 1.4.1). Leerlingen met een gebrekkige decodeervaardigheid zullen tevens beperkt worden in het begrijpen van een tekst. Uit de grootte van de parameterschattingen kunnen we afleiden dat de invloed van een toename in technische leesvaardigheid het grootst is tussen de eerste twee perioden.

Helemaal aan het begin, tussen het voorjaar van groep 3 en het najaar van groep 4, zien we tevens een invloed van begrijpend lezen op technisch lezen gemeten met de AVI-



Figuur 4.7: Uiteindelijk model voor de longitudinale relatie tussen AVI en Begrijpend Lezen

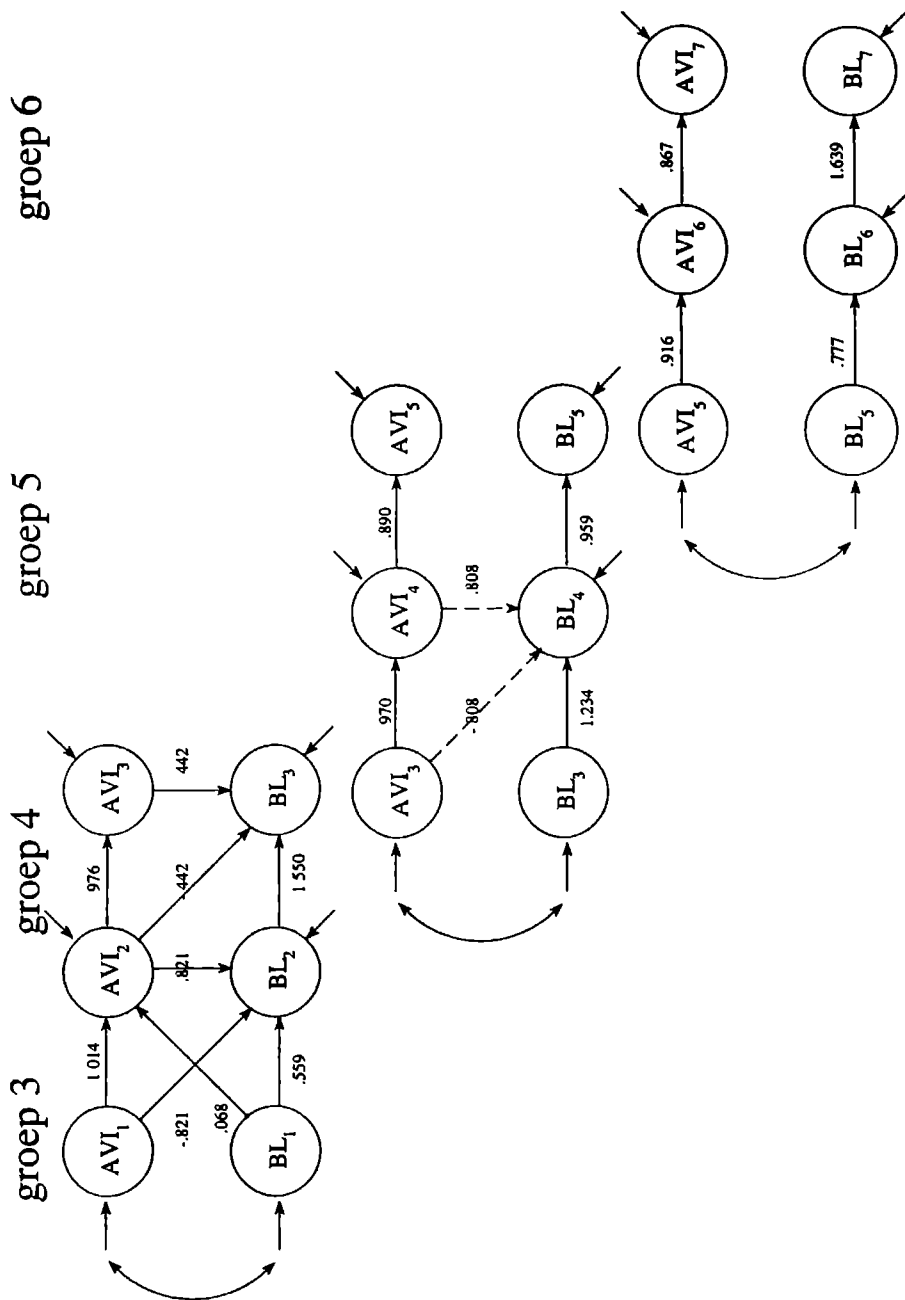
toetskaarten. Dit effect is vanuit de theorie moeilijk te verklaren. Een mogelijke verklaring zou kunnen zijn dat beginnende lezers hun nog gebrekkige technische leesvaardigheid kunnen compenseren door hun vaardigheid in begrijpend lezen. Doordat de AVI-toetskaarten bestaan uit teksten die een afgerond verhaal vormen, kunnen de beginnende lezers gebruik maken van de context (en dus hun vaardigheid in begrijpend lezen) bij het decoderen van de tekst (zie paragraaf 1.2.3). Toetsing van dezelfde modellen voor de relatie tussen technisch lezen gemeten met de Eén-Minuuut-Test en begrijpend lezen geeft echter hetzelfde eindmodel. Daarmee vervalt deze verklaring, aangezien de context geen rol kan spelen bij de Eén-Minuuut-Test.

Wellicht dient de verklaring gezocht te worden in het volgende. Voor de AVI-toets geldt dat in het voorjaar van groep 3 sprake is van een bodem-effect (zie tabel 4.3). Een groot deel van de leerlingen beheerst AVI-niveau 1 nog niet (bijna 45%, zie tabel 3.14 in hoofdstuk 3). Dit betekent dat het onderscheidend vermogen van de toets tussen zwakkere en minder zwakke lezers beperkt is. Voor de toetsen voor begrijpend lezen in het voorjaar van groep 3 geldt dat juist eerder sprake is van een plafondeffect. Deze toetsen maken dus wel voldoende onderscheid tussen de zwakkere en minder zwakke lezers, maar niet tussen goede en iets minder goede lezers. Aangezien de vaardigheden in technisch lezen en begrijpend lezen zeker bij beginnende lezers sterk samenhangen, betekent dit dat de toetsen voor begrijpend lezen tevens in enige mate onderscheid maken in zwakke en minder zwakke technische lezers. Het gebrek aan onderscheidend vermogen van de AVI-toets in het voorjaar van groep 3 kan door de toetsen voor begrijpend lezen als het ware gecompenseerd worden, hetgeen kan verklaren waarom juist tussen het voorjaar van groep 3 en het najaar van groep 4 sprake is van een effect van begrijpend lezen op technisch lezen.

Het tweede deel van de figuur betreft model C3 van cohort 4. Tussen het voorjaar van groep 4 en het najaar van groep 5 is nog steeds sprake van een invloed van een verbetering in technisch lezen op het begrijpend lezen. We hebben deze invloed echter weergegeven met een stippellijn, omdat uit de toetsingsresultaten bleek dat er slechts sprake is van een zwakke aanwijzing voor het bestaan ervan.

Vanaf het najaar van groep 5 is er geen sprake meer van een invloed van een toename in de decodeervaardigheid op begrijpend lezen. Deze trend zien we heel duidelijk in het laatste deel van de figuur (model D1 van cohort 5). Er is wel sprake van een samenhang tussen de prestaties in technisch en begrijpend lezen, hetgeen wordt weergegeven door de correlatie aan het begin van de longitudinale reeks. Het is echter niet meer zo dat veranderingen in technisch lezen nog een invloed hebben op het niveau van begrijpend lezen. Deze resultaten bevestigen de uitkomsten van ander onderzoek dat prestaties en vorderingen in technisch lezen vooral in het begin van het basisonderwijs belangrijk zijn voor het begrijpen van teksten. De sterkte van deze invloed neemt af naarmate het onderwijs vordert.

In figuur 4.8 hebben we de uiteindelijke longitudinale modellen weergegeven voor de relatie tussen AVI en Woordenschat. Daarin hebben we voor de structurele relaties tussen de latente variabelen de ongestandaardiseerde parameterschattingen opgenomen. In bijlage 18 staan alle ongestandaardiseerde parameterschattingen weergegeven voor de betreffende modellen. Aangezien de uitkomsten overeenkomen met die voor de relatie tus-



Figuur 4.8: Uiteindelijk model voor de longitudinale relatie tussen AVI en Woordenschat

sen AVI en Begrijpend Lezen zullen we deze figuur niet nader bespreken. Wel willen we verwijzen naar de modeluitkomsten in paragraaf 4.4.4 waaruit blijkt dat de resultaten voor de relatie tussen AVI en woordenschat niet zo eenduidig zijn als die voor de relatie tussen AVI en begrijpend lezen.

4.5 Conclusies

In dit onderzoek hebben we de validiteit van de AVI-toetskaarten onderzocht, uitgaande van de onderzoeksvragen 4 tot en met 6. Daarbij hebben we gebruik gemaakt van Lisrel-modellen.

Uit het onderzoek blijkt dat de twee versies van de AVI-toetskaarten als paralleltests beschouwd mogen worden (onderzoeksvraag 4). De gemiddelden op beide versies zijn gelijk, evenals de factorladingen op de latente factor AVI en de errorvarianties. De geschatte betrouwbaarheid van de AVI-toetskaarten ligt tussen .85 en .91.

Uit de correlaties tussen de latente factoren AVI en EMT blijkt dat deze sterk samenhangen (correlaties tussen 0,81 en 0,90). Aangezien beide toetsen bedoeld zijn voor het meten van de technische leesvaardigheid is dit een bevestiging van de validiteit van de AVI-toetskaarten. De correlaties tussen beide factoren konden echter niet gelijk gesteld worden aan 1 of 0,95. Dit betekent dat de AVI-toetskaarten en de Eén-Minuut-Test niet precies hetzelfde aspect meten van technisch lezen (onderzoeksvraag 5).

De relatie tussen de AVI-toetskaarten en Begrijpend Lezen en Woordenschat hebben we eerst per cohort per periode onderzocht en vervolgens longitudinaal. Per cohort per periode is de correlatie geschat tussen de latente AVI-factor en de latente variabelen Begrijpend Lezen en Woordenschat. Uit deze voor onbetrouwbaarheid gecorrigeerde correlaties blijkt dat de AVI-toetskaarten duidelijk iets anders meten dan de toetsen voor begrijpend lezen en woordenschat. De correlaties tussen de AVI-factor en deze twee variabelen liggen beduidend lager dan de correlaties van de AVI-toetskaarten met de Eén-Minuut-Test.

De samenhang tussen de AVI-factor enerzijds en Begrijpend Lezen en Woordenschat anderzijds is het hoogst in groep 4, en neemt af in groep 5 en groep 6. Dit komt overeen met wat we op basis van de theorie mochten verwachten: de samenhang tussen technisch lezen enerzijds en begrijpend lezen en woordenschat anderzijds is het grootst aan het begin van het leesonderwijs.

Tot slot blijkt uit de longitudinale analyses dat met name in de eerste jaren waarin een kind leert lezen een toename in de technische leesvaardigheid een effect heeft op de vaardigheid in begrijpend lezen. Tussen het voorjaar van groep 3 en het najaar van groep 4, tussen het najaar van groep 4 en het voorjaar van groep 4 en tussen het voorjaar van groep 4 en het najaar van groep 5 heeft een toename in het niveau van technisch lezen invloed op het begrijpend lezen van een leerling. Deze invloed neemt al wel af gedurende deze perioden.

Tussen het voorjaar van groep 3 en het najaar van groep 4, de periode waarin de meeste leerlingen de basis van het technisch lezen onder de knie krijgen, zien we ook nog de

invloed van begrijpend lezen op technisch lezen zoals gemeten met AVI. Het AVI-niveau in najaar van groep 4 wordt mede bepaald door het niveau van begrijpend lezen in groep 3. Dit is goed te verklaren door de vorm van de AVI-toets: teksten met een afgerond verhaal. Daardoor kunnen de beginnende lezers gebruik maken van de context en dus hun vaardigheid in begrijpend lezen bij het hardop lezen van de tekst. Uit de literatuur is bekend dat beginnende lezers gebruik maken van de context bij het decoderen. Vanaf het najaar van groep 5 is geen sprake meer van een invloed van (een winst in) technisch lezen op begrijpend lezen. Uiteraard is er wel sprake van een samenhang tussen de prestaties in technisch en begrijpend lezen; voor cohort 5 hebben we dan ook een correlatie opgenomen tussen de factoren AVI en Begrijpend Lezen aan het begin van de longitudinale reeks. Het is echter niet meer zo dat veranderingen in technisch lezen nog een invloed hebben op het niveau van begrijpend lezen.

5 DE KWALITEIT VAN DE NIEUWE AVI-TOETSKAARTEN

De AVI-toetskaarten vormden een onderdeel van het project 'Analyse Van Individualiseringsvormen' (AVI) van het KPC uit 1972 (zie paragraaf 1.1.1). Dit project was bedoeld om hulpmiddelen en werkwijzen te ontwikkelen voor de individualisering van het gehele leesonderwijs in de basisschool, zowel voor de leestechiek, het leesbegrip als de leesbeleving. Speciaal voor het technisch leesonderwijs zijn negen AVI-niveaus gecreëerd. De AVI-niveaus 1 tot en met 4 variëren in moeilijkheidsgraad van éénlettergrepige woorden tot en met drie- en meerlettergrepige woorden, de hogere niveaus zijn uitsluitend ingedeeld op basis van de Leesindex A van Brouwer, waarin de woord- en zinslengte de bepalende factoren zijn. Deze niveaus worden enerzijds gebruikt om leesboeken in te delen naar moeilijkheidsgraad. Anderzijds worden de prestaties op technisch lezen van de leerlingen ingedeeld naar deze niveaus. Voor dit laatste zijn de AVI-toetskaarten ontwikkeld. Het doel van de AVI-toetskaarten is het vaststellen van het niveau van technisch lezen van kinderen om zo, ten behoeve van het technisch leesonderwijs, leesstof te kunnen kiezen die aansluit bij hun niveau.

In dit proefschrift hebben we onze aandacht gericht op de AVI-toetskaarten: hoe betrouwbaar en valide zijn deze om de technische leesvaardigheid van de leerlingen vast te stellen. We hebben daarbij de relatie onderzocht tussen de toetskaarten en de technische leesvaardigheid van de leerlingen. Op een continuüm van technische leesvaardigheid hebben we zowel de positie van de toetskaarten als de positie van de leerlingen bepaald. De relatie tussen de leesstof en dit continuüm viel buiten het bestek van dit proefschrift. In dit laatste hoofdstuk geven we in paragraaf 5.1 een terugblik op de eerste vier hoofdstukken. Daarbij zullen we tevens per onderzoeksvraag de conclusies weergeven. Vervolgens bespreken we in paragraaf 5.2 de kwaliteit van de nieuwe toetskaarten, de waarde van het onderzoek en de toetskaarten voor de onderwijspraktijk en mogelijke wegen voor verder onderzoek.

5.1 Samenvatting en conclusies

5.1.1 De AVI-toetskaarten

De AVI-toets bestaat uit negen kaarten, voor elk AVI-niveau één. Een AVI-niveau verwijst naar een bepaalde technische moeilijkheidsgraad van teksten. Op elke kaart staat een tekst afgebeeld, welke een afgerond verhaal vormt. De teksten op de kaarten voldoen telkens aan de criteria van het niveau waarvoor ze bedoeld zijn. Van de AVI-toets zijn een A- en een B-versie beschikbaar.

De afname van de toets vindt individueel plaats. De leerling krijgt een kaart voor zich die behoort bij het niveau waarop hij of zij geacht wordt te lezen. De leerling moet de tekst hardop voorlezen. Het aantal fouten dat de leerling maakt en de tijd (in seconden) die hij of zij nodig heeft om de tekst te lezen bepalen de leesprestatie, ofwel het AVI-niveau, van de leerling. Wanneer de prestatie van de leerling *voldoende* is, kan worden gestopt met de afname en wordt de leerling ingedeeld op het betreffende AVI-niveau. Bij een *onvoldoende* prestatie krijgt de leerling de kaart van een niveau lager voorgelegd, bij een *goede* leesprestatie de kaart van een niveau hoger. De toetsafname gaat op deze

wijze verder totdat de prestatie *voldoende* is bereikt. Op deze manier kan een leerkracht snel en eenvoudig te weten komen hoe goed de leerlingen lezen en kan de leerkracht vervolgens, met behulp van de niveaulijsten (waarin leesboeken zijn geordend naar AVI-niveau), de leerlingen die stof laten lezen die afgestemd is op hun technische leesvaardigheid.

In het proces van de verwerving van de technische leesvaardigheid onderscheiden we drie fasen (zie paragraaf 1.3.1). In de eerste fase herkent een kind woorden op basis van willekeurige kenmerken als bijvoorbeeld de beginletter. In de tweede fase maakt het kind expliciet gebruik van de letter-klankkoppeling bij het identificeren van woorden. Tot slot wordt in de derde fase de decodeervaardigheid geautomatiseerd. De AVI-toetskaarten zijn bedoeld voor dit laatste stadium, hetgeen betekent dat de toets moet bestaan uit woorden, zinnen of teksten. Letters, maar ook pseudowoorden zijn niet geschikt als testmateriaal. Immers een toets die is opgebouwd uit pseudowoorden meet expliciet de beheersing van de grafeem-foneemkoppeling en is daarmee beperkt tot het meten van het beheersen van het eerder genoemde tweede stadium van de leesontwikkeling.

Voor de nieuwe AVI-toets is, net als bij de oude, gekozen voor teksten als toetsmateriaal. Een eerste reden om daarvoor te kiezen was dat de toets in die vorm breed geaccepteerd is in het onderwijs. Een tweede reden was de ecologische validiteit: door gebruik te maken van teksten lijkt de testsituatie meer op de werkelijke leessituatie, waarin men eveneens samenhangende teksten leest. Een volgende reden was dat de AVI-toets met teksten als toetsmateriaal een aanvulling vormt op de bestaande woordentests voor technisch lezen (Eén-Minuut-Test en de Drie-Minuten-Toets). Het idee overigens achter het gebruiken van samenhangende teksten is dat de context een rol kan spelen bij het technisch lezen. Met name beginnende lezers maken gebruik van de context bij woordherkenning. Gebleken is dat de context bij de ervaren lezers niet of nauwelijks een rol speelt bij woordherkenning. Naarmate de decodeervaardigheid meer geautomatiseerd wordt, neemt de invloed van de context af. Het gegeven dat kinderen wel gebruikmaken van de context bij het decoderen was een laatste reden om gehele teksten als toetsmateriaal te gebruiken.

Zoals al eerder is gezegd is de AVI-toets bedoeld om vast te stellen in hoeverre de technische leesvaardigheid geautomatiseerd is. Geautomatiseerd wil zeggen dat een woord nauwkeurig, 'automatisch' (zonder bewuste aandacht) en vooral snel herkend wordt. De snelheid van woordherkenning is daarbij de beste indicator voor de mate waarin de vaardigheid is geautomatiseerd. Met behulp van de AVI-toets wordt zowel de nauwkeurigheid (het aantal fouten) als de snelheid (benodigde leestijd) van de decodeervaardigheid vastgesteld.

De maximale snelheid van decoderen wordt overigens niet voor alle woorden tegelijkertijd bereikt: korte, regelmatig gespelde en frequent voorkomende woorden worden eerder 'automatisch' en snel gelezen dan lange, onregelmatig gespelde en minder frequent voorkomende woorden. In de diverse methoden voor aanvankelijk leesonderwijs zien we een overeenkomstige opbouw in moeilijkheidsgraad van woorden die worden aangeleerd: men start met eenvoudige éénlettergrepige woorden. De AVI-toets is daarom opgebouwd uit

teksten van oplopende moeilijkheidsgraad waarbij voor de niveaus 1 tot en met 4 tevens rekening is gehouden met de opbouw van de methoden voor aanvankelijk lezen.

5.1.2 Herziening van de AVI-toetskaarten

In de periode 1992-1994 zijn de nieuwe AVI-toetskaarten ontwikkeld. De eerste reden voor de herziening was het feit dat de teksten en illustraties - ontwikkeld in de jaren zeventig - niet meer afgestemd waren op de belevingswereld van de kinderen in de huidige tijd. Een tweede reden om de toets te herzien betrof de criteria voor de opbouw in moeilijkheidsgraad van de eerste vier AVI-niveaus. Deze waren vrijwel uitsluitend gebaseerd op de oude versie van de methode Veilig Leren Lezen. Gezien de vele nieuwe beschikbare methoden voor aanvankelijk lezen diende deze indeling opnieuw bekeken te worden. Een derde reden was de normering van de toetskaarten. Gegevens over de totstandkoming en de representativiteit daarvan ontbraken, evenals gegevens over de validiteit van de toetskaarten. Onderzoek van Dickhout-Rutten (1983) wees uit dat de oorspronkelijke indelingscriteria naar AVI-niveau op grond van leestijd en aantal gemaakte fouten niet voldeden (zie paragraaf 1.1.3). Uit onderzoek van Touwen (1990) bleek tevens dat meer dan de helft van de scholen andere criteria voor tijd en fouten hanteerden (veelal afkomstig van schoolbegeleidingsdiensten) dan door het KPC werden voorgeschreven. Deze praktijk leidde ertoe dat leerlingen met exact dezelfde prestaties op een toetskaart ingedeeld werden in verschillende AVI-niveaus, afhankelijk van de gebruikte criteria.

De constructie van de nieuwe AVI-toetskaarten is verlopen in een aantal fasen (zie paragraaf 2.2). In de eerste fase zijn de methode-afhankelijke criteria verwijderd uit de eerste vier AVI-niveaus; voor de overige vijf niveaus is de indeling gelijk gebleven. Vervolgens zijn nieuwe teksten geschreven voor de negen AVI-niveaus waarbij getracht is zoveel mogelijk aan te sluiten bij de belevingswereld van kinderen van nu. Deze teksten zijn ter beoordeling voorgelegd aan een aantal leesdeskundigen. Hun commentaar is door een bureauredactrice verwerkt, rekening houdend met het feit dat de kaarten moesten voldoen aan de criteria van de AVI-niveaus. De resulterende teksten zijn vervolgens uitgeprobeerd op een school en tevens door 25 leerkrachten beoordeeld op moeilijkheidsgraad. Op basis van deze eerste try-out werden de teksten nogmaals aangepast. Deze aangepaste teksten zijn vervolgens uitgeprobeerd op zeven basisscholen. Vrij snel na deze tweede try-out bleek dat voor de schaling, normering en validering van de toetskaarten kon worden aangehaakt bij het longitudinale onderzoek naar schoolvorderingen van de Vakgroep Onderwijskunde Nijmegen. Doordat dit onderzoek reeds begonnen was en de AVI-toetskaarten daarin op korte termijn afgenomen moesten worden, was er geen tijd meer beschikbaar om nadere analyses uit te voeren op de gegevens van de tweede try-out en op basis daarvan de toetskaarten zo nodig nog te wijzigen.

De volgende onderzoeksvragen stonden in het onderzoek, waarvan in dit proefschrift verslag is gedaan, centraal:

- 1 Lopen de nieuwe AVI-toetskaarten op in moeilijkheidsgraad?
- 2 Wat is de betrouwbaarheid van de AVI-toetskaarten?
- 3 Hoe is de verdeling van de AVI-niveaus over de basisschoolperiode?

- 4 Kunnen de A- en de B-versie van de AVI-toets als paralleltests worden beschouwd?
- 5 Wat is de relatie tussen de AVI-toets en de Eén-Minuut-Test (EMT)?
- 6 Wat is de relatie tussen de AVI-toets en toetsen voor begrijpend lezen en woordenschat?

De resultaten betreffende de eerste drie onderzoeksvragen staan weergegeven in paragraaf 5.1.3, de overige vragen worden besproken in paragraaf 5.1.4.

5.1.3 De schaling en normering

Steekproef

Ten behoeve van de schaling en normering van de AVI-toetskaarten zijn deze kaarten afgenomen bij een steekproef van leerlingen uit de doelgroep: leerlingen van groep 3 tot en met 6 van het basisonderwijs. De steekproef is afkomstig uit het longitudinale onderzoek van de Vakgroep Onderwijskunde Nijmegen waarin drie cohorten leerlingen worden gevolgd gedurende de basisschoolperiode: leerlingen die in september 1991 aanwezig waren in respectievelijk groep 3, 4 en 5. Overigens is deze steekproef niet volledig representatief voor alle Nederlandse basisscholen: scholen in grote steden met veel allochtone leerlingen en Nederlandse achterstandsleerlingen zijn ondervertegenwoordigd.

Voor de schaling en normering van de AVI-toetskaarten is gebruik gemaakt van de gegevens verzameld in maart 1992, oktober 1992 en maart 1993. In maart 1992 beschikten we over gegevens van 1049 leerlingen in groep 3, 1029 leerlingen in groep 4 en 930 leerlingen in groep 5. In oktober 1992 betrof het de gegevens van 950 leerlingen in groep 4, 850 leerlingen in groep 5 en 805 leerlingen in groep 6 en in maart 1993 waren gegevens beschikbaar van 942 leerlingen in groep 4, 849 in groep 5 en 804 in groep 6.

Dataverzameling

In maart 1992 en maart 1993 werd bij alle leerlingen zowel de A- als de B-versie van de nieuwe AVI-toets afgenomen, in oktober 1992 alleen de A-versie. In principe kregen de leerlingen drie opeenvolgende toetskaarten van een versie te lezen die qua moeilijkheid in de buurt van hun leesniveau lagen. Dit leesniveau is in maart 1992 ingeschat met behulp van de Eén-Minuut-Test voor de leerlingen van groep 4 en 5 en met Caesar-EMT voor de leerlingen van groep 3. In oktober 1992 en maart 1993 is dit leesniveau ingeschat op basis van de prestaties op de nieuwe AVI-toetskaarten in de voorgaande periode. In het onderzoek zijn de benodigde leestijd en het aantal fouten per toetskaart gebruikt.

Bepalen van de grenswaarden

De eerste stap in het onderzoek was het bepalen van de grenswaarden: bij welk aantal fouten en welke leestijd behoort welk AVI-niveau? Per kaart worden drie prestatieniveaus onderscheiden: onvoldoende, voldoende en goed. Voor zowel de leestijd als het aantal fouten dienden daarom twee grenswaarden vastgesteld te worden, namelijk voor de overgang van onvoldoende naar voldoende en voor de overgang van voldoende naar goed. Die grenswaarden dienden zodanig gekozen te worden dat de uitkomsten van de verschillende kaarten zo consistent mogelijk waren. Om dit te bewerkstelligen is de volgende aanpak gevolgd.

Voor de schaling van de toetskaarten hebben we gebruik gemaakt van het door Jansen (1997) bewerkte model voor leestijden van Rasch (1980) (zie paragraaf 3.2.2). Een model waarin zowel de leestijd als het aantal fouten betrokken konden worden was niet beschikbaar. Gezien het feit dat de leessnelheid de beste indicator is voor de mate waarin de decodeervaardigheid is geautomatiseerd, lag het voor de hand de leestijd als uitgangspunt te nemen.

Nadat de AVI-toetskaarten op basis van de leestijden onderling waren geschaald is de resulterende schaal voor de technische leesvaardigheid opgedeeld in de negen AVI-niveaus. De bijbehorende grenswaarden voor de leestijden zijn zodanig gekozen dat de resultaten op de verschillende toetskaarten zoveel mogelijk met elkaar overeen kwamen. Ten behoeve van de hanteerbaarheid in de praktijk zijn de leestijden afgerond op veelvouden van vijf seconden. Vervolgens zijn de grenswaarden voor de leesfouten vastgesteld uitgaande van het principe dat de leessnelheid niet ten koste mocht gaan van de nauwkeurigheid. Als uitgangspunt voor de grenswaarden van de voldoende-grens is bij iedere toetskaart het gemiddelde aantal fouten genomen plus 1 standaardafwijking van degenen die op basis van de leestijden in het AVI-niveau van de betreffende kaart zijn ingedeeld. Voor de goed-grens is als uitgangspunt gekozen het gemiddelde aantal fouten plus 1 standaardafwijking van degenen die volgens de leestijden minimaal één AVI-niveau hoger beheersen dan de betreffende kaart.

Onderzoeksresultaten

1 Lopen de AVI-toetskaarten op in moeilijkheidsgraad?

Tijdens de constructiefase is ervoor gezorgd dat de teksten op de toetskaarten voldoen aan de criteria behorende bij het niveau waarvoor de kaart bedoeld is. Toepassing van het model van Rasch zoals aangepast door Jansen leverde schattingen op van de moeilijkheidsgraad van de teksten, waarbij de leessnelheid als uitgangspunt is genomen (zie paragraaf 3.2.2). Het verloop van deze geschatte moeilijkheidsgraden was niet altijd conform de veronderstelde moeilijkheidsgraden van de kaarten. Zo is in versie A kaart 4A moeilijker dan de kaarten 5A en 6A. Voor versie B geldt dat kaart 4B makkelijker is dan kaart 3B, kaart 6B makkelijker is dan kaart 5B, terwijl de kaarten 7B en 8B even moeilijk zijn. Ook verschilde de moeilijkheid van kaarten die bedoeld waren voor hetzelfde niveau nogal eens. Voor een aantal AVI-toetskaarten lagen de geschatte moeilijkheidsgraden dicht bij elkaar en waren de verschillen in moeilijkheid nogal klein. Zo lagen de moeilijkheidsgraden van de kaarten 3A, 4A, 5A, 5B, 6A en 6B vrij dicht opeen, evenals de moeilijkheidsgraden van de kaarten voor de niveaus 7, 8 en 9. We moeten daarom concluderen dat de volgorde van de AVI-toetskaarten naar moeilijkheidsgraad niet volledig overeenkomt met de ordening naar AVI-niveau.

Bij het bepalen van de grenswaarden voor de leestijden per kaart hebben we rekening gehouden met de problemen ten aanzien van de moeilijkheidsgraad (zie paragraaf 3.2.3). Zo zijn de grenswaarden voor de leestijden bij relatief te makkelijke kaarten strenger dan voor de relatief moeilijkere teksten. Een gevolg hiervan was wel dat voor de kaarten vanaf niveau 5 het verschil tussen de leestijden voor de voldoende- en de goed-grens klein is (10 tot 15 seconden), waardoor kleine verschillen in leestijd kunnen leiden tot een verschil van één of twee AVI-niveaus.

2 Wat is de betrouwbaarheid van de AVI-toetskaarten?

Bij de procedure voor het bepalen van de grenswaarden voor de toetskaarten hebben we getracht deze waarden zo te kiezen dat de resultaten op de verschillende toetskaarten zo consistent mogelijk waren. Op drie manieren is vervolgens nagegaan in hoeverre dit het geval was (zie paragraaf 3.3).

Op de eerste plaats hebben we gecontroleerd of de uitkomsten bij twee kaarten van hetzelfde niveau (versie A versus versie B) hetzelfde zijn: als de uitkomst van een kaart van versie A onvoldoende (respectievelijk voldoende of goed) is, is de uitkomst van de kaart van hetzelfde niveau van versie B dan eveneens onvoldoende (respectievelijk voldoende of goed)? De vergelijkbaarheid van de resultaten op kaartniveau was matig tot redelijk: het percentage consistente uitkomsten lag tussen de 62% en 82%. De consistentie was het hoogst voor de niveaus 1, 2 en 3 (75%-82%). Voor de niveaus 4, 8 en 9 was 69%-70% van de uitkomsten gelijk. Het laagst was de consistentie voor de niveaus 5, 6 en 7 (62%-64%).

Vervolgens is nagegaan of de uitkomsten van twee opeenvolgende kaarten van een versie overeenkwamen. Met uitzondering van de vergelijking van de kaarten 6B en 7B (68% consistent) was het percentage consistente uitkomsten 75% of hoger. De consistentie was het hoogst voor de vergelijking tussen de niveaus 1, 2, 3 en 4 (81%-89%) en tussen de niveaus 8 en 9 (80%-81%). De grootste inconsistenties vonden we bij de kaarten die qua moeilijkheidsgraad dicht bij elkaar lagen of waarvan de empirische volgorde omgekeerd was aan de volgorde naar AVI-niveau. In de meeste gevallen impliceerde een inconsistentie uiteindelijk een verschil van één AVI-niveau.

Tot slot hebben we de overeenkomst bepaald tussen de uiteindelijke uitkomsten van beide versies van de AVI-toets. De overeenkomst tussen het uiteindelijke AVI-niveau volgens versie A en B was enigszins teleurstellend: 62,1%. Als de niveaus niet overeenkwamen betrof het meestal een verschil van één AVI-niveau.

Naast de interne consistentie van de nieuwe toets is ook de overeenstemming met de oude toets bepaald, gebruikmakend van de gegevens uit de tweede try-out. Deze overeenstemming bleek niet groot te zijn: 41% tussen versie A van de oude toets en de versies A en B van de nieuwe toets. Over het algemeen kwam gemiddeld genomen het niveau volgens de nieuwe kaarten wat lager uit dan volgens de oude versie A. Het was ook niet de bedoeling om volledige overeenstemming te krijgen, aangezien de verdeling van de niveaus volgens de oude toets te wensen overliet en een van de redenen tot herziening was. Wel heeft deze matige overeenstemming gevolgen voor de praktijk, wanneer men overstapt van de oude naar de nieuwe toetskaarten. Het kan voorkomen dat het AVI-niveau van de leerling volgens de nieuwe AVI-toets anders is dan verwacht (zie ook de handleiding van de AVI-toets van Visser, Van Laarhoven & Ter Beek, 1996).

3 Hoe is de verdeling van de AVI-niveaus over de basisschoolperiode?

Toepassing van het aangepaste Rasch-model ten behoeve van de schaling van de toetskaarten leverde ook schattingen op van de technische leesvaardigheid van de leerlingen (zie paragraaf 3.2.2). Het verloop van deze schattingen kwam overeen met wat we op basis van de theorie mochten verwachten. Tussen maart van groep 3 en maart van groep 6 zagen we een duidelijke toename van de technische leesvaardigheid. Deze groei in

technische leesvaardigheid was het sterkst aan het begin van het leesonderwijs en nam af in de hogere jaargroepen. Een vergelijkbare ontwikkeling was zichtbaar bij de AVI-niveaus (zie paragraaf 3.4). In de loop van de basisschoolperiode nam het AVI-niveau van de leerlingen duidelijk toe. In maart van groep 3 beheerste 44% van de leerlingen AVI-niveau 1 nog niet. Uiteindelijk beheerste circa 50% van de leerlingen AVI-niveau 9 of hoger in maart van groep 6. De verschuiving van de AVI-niveaus in de tijd verliep gelijkmatig, zonder grote sprongsgewijze verschuivingen. Vergelijking van het gemiddelde AVI-niveau in de tijd liet zien dat de groei aan het begin van het leesonderwijs het sterkst was: van gemiddeld 1,1 in maart van groep 3 naar gemiddeld 4,5 in maart van groep 4, een groei van meer dan drie niveaus. Verderop in het basisonderwijs was deze groei minder sterk: van 6,9 in maart van groep 5 naar 8,3 in maart van groep 6, een groei van ruim één AVI-niveau.

5.1.4 De validering

De data

Voor het validiteitsonderzoek zijn de gegevens gebruikt van dezelfde steekproef leerlingen in dezelfde perioden (voorjaar 1992, najaar 1992 en voorjaar 1993) als voor de schaling en normering van de toetskaarten. Van deze leerlingen zijn de toetsgegevens gebruikt op de terreinen technisch lezen, begrijpend lezen en woordenschat. Een probleem hierbij was het ontbreken van toetsgegevens. Toetsgegevens ontbraken doordat sommige leerlingen één of meer toetsen in een bepaalde periode niet gemaakt hadden. Bovendien ontbraken van een leerling soms alle gegevens in een periode doordat de leerling was blijven zitten, verhuisde of nieuw op school was gekomen of doordat een school stopte met deelname aan het onderzoek. Om maximaal gebruik te kunnen maken van de beschikbare gegevens zijn de ontbrekende gegevens vervangen met behulp van de imputatie-procedure uit PRELIS 2 (zie paragraaf 4.1.1). Dit is gebeurd per cohort per periode, onder de beperkende voorwaarde dat voor een leerling minimaal vier toetsgegevens in de betreffende periode bekend moesten zijn. Leerlingen met te veel ontbrekende gegevens zijn niet in de analyses betrokken. De uiteindelijke uitval van leerlingen per cohort per periode bedroeg 10% tot 15% in maart 1992 en circa 1% in oktober 1992 en maart 1993. De longitudinale uitval tussen maart 1992 en maart 1993 bedroeg circa 23%. Uit nader onderzoek bleek dat sprake was van selectieve uitval van leerlingen; de gegevens zijn niet 'missing at random'. Met name voor cohort 3 gold dat de leerlingen die in het validiteitsonderzoek werden betrokken betere prestaties leverden op de toetsen voor technisch lezen dan de andere leerlingen.

Werkwijze

Ter beantwoording van de onderzoeksvragen 4 tot en met 6 is gebruik gemaakt van Lisrel-modellen (Lisrel 8). De gevolgde strategie betrof de volgende stappen: eerst de specificatie van het basismodel, vervolgens uitbreiding dan wel beperking van het model conform de te toetsen hypothesen, daarna de hypothese-toetsing en tot slot de parameterschatting en interpretatie (zie paragraaf 4.1.2).

Onderzoeksresultaten

4 Kunnen de A- en B-versie van de AVI-toets als paralleltests worden beschouwd?

Voor de beantwoording van deze vraag zijn Lisrel-modellen toegepast op de gegevens per cohort per periode (zie paragraaf 4.2.2). De resultaten hiervan geven aan dat we de twee versies van de AVI-toetskaarten als paralleltests mogen beschouwen. De gemiddelden op beide versies zijn gelijk, alsmede de factorladingen op de latente variabele en de error-varianties.

5 Wat is de relatie tussen de AVI-toets en de EMT?

De correlaties (gecorrigeerd voor onbetrouwbaarheid) tussen de latente variabelen AVI en EMT, beide gemeten met behulp van de A- en B-versie van de betreffende toetsen, waren hoog: tussen 0,85 en 0,90 voor groep 3 tot en met 5, en tussen 0,81 en 0,83 voor groep 6. De correlatie lijkt redelijk stabiel in de groepen 4 en 5, maar neemt iets af in groep 6. Het plafond-effect bij de AVI-toetskaarten kan daarvan een oorzaak zijn.

De correlaties zijn weliswaar hoog, maar niet perfect (statistisch significant lager dan 0,95), hetgeen betekent dat de toetsen niet precies hetzelfde meten. Door gebruik te maken van teksten in plaats van woorden als toetsmateriaal wordt met de AVI-toets net een ander aspect van de technische leesvaardigheid gemeten dan met de Eén-Minuut-Test.

6 Wat is de relatie tussen de AVI-toets en toetsen voor begrijpend lezen en woordenschat?

De relatie tussen de AVI-toetskaarten en Begrijpend Lezen en Woordenschat hebben we eerst per periode en vervolgens longitudinaal onderzocht. Per cohort per periode is nagegaan wat de relatie is tussen de AVI-toets enerzijds en toetsen voor begrijpend lezen en woordenschat anderzijds. De geschatte correlaties tussen de latente variabelen AVI en Begrijpend lezen waren over het algemeen van dezelfde orde van grootte als die tussen AVI en Woordenschat. De correlaties waren het hoogst in groep 4 (0,60-0,70) en namen af in groep 5 (0,55-0,60) en 6 (0,46-0,55). De correlaties waren beduidend lager dan die tussen AVI en EMT, conform de verwachting. De AVI-toetskaarten meten duidelijk iets anders dan toetsen voor begrijpend lezen en woordenschat.

Vervolgens is voor ieder cohort een longitudinaal model ontwikkeld voor de relatie tussen AVI en Begrijpend lezen respectievelijk Woordenschat. De resultaten gaven aan dat in de eerste jaren waarin een kind leert lezen een winst in de technische leesvaardigheid een positief effect heeft op het begrijpend lezen en de woordenschat. Een toename in technische leesvaardigheid heeft een positieve invloed op de prestaties bij begrijpend lezen en op de woordenschat. Deze invloed neemt af naarmate de (basisschool)tijd vordert. Vanaf het najaar van groep 5 hangen beide vaardigheden wel duidelijk met elkaar samen (correlatie van circa 0,50), maar heeft een verdere toename van de technische leesvaardigheid (gemeten met de AVI-toetskaarten) geen extra effect meer op de prestaties in begrijpend lezen. Anders gezegd: veranderingen in de technische leesvaardigheid hebben in de latere leerjaren geen invloed meer op het niveau van begrijpend lezen.

5.2 Discussie

In maart 1994 zijn de nieuwe AVI-toetskaarten op de markt gebracht ter vervanging en verbetering van de oude kaarten uit 1977. In deze paragraaf staan we stil bij de kwaliteit van de nieuwe toetskaarten. We beginnen in paragraaf 5.2.1 met de psychometrische kwaliteit: hoe betrouwbaar en valide is de nieuwe AVI-toets. Vervolgens gaan we in paragraaf 5.2.2 in op de waarde van de AVI-toetskaarten voor de praktijk. In paragraaf 5.2.3 schetsen we de bijdrage van het onderzoek naar de nieuwe AVI-toetskaarten aan de theorievorming rond technisch lezen en geven we enkele suggesties voor vervolgonderzoek rond technisch lezen en de AVI-toetskaarten in het bijzonder.

5.2.1 De psychometrische kwaliteit

Het onderzoek in dit proefschrift was geconcentreerd rond de psychometrische kwaliteit van de AVI-toetskaarten. Ter beantwoording van de vraag hoe het met deze kwaliteit is gesteld gaan we in op drie aspecten: de opbouw in moeilijkheidsgraad van de negen AVI-niveaus, de mate waarin de AVI-toets de technische leesvaardigheid op een betrouwbare manier meet en de mate waarin de toets meet wat we ermee willen meten (de validiteit).

De basis van de AVI-toetskaarten wordt gevormd door de negen AVI-niveaus, die een toenemende mate van zowel de technische leesvaardigheid van de leerlingen als de technische moeilijkheidsgraad van de teksten beogen te weerspiegelen. In het onderzoek hebben we zowel de leerlingen als de toetskaarten geschaald op een onderliggend continuüm van technische leesvaardigheid. Uit de schaling is gebleken dat de volgorde van de toetskaarten naar moeilijkheidsgraad (op basis van de leestijden) niet geheel overeenkwam met de ordening op basis van de AVI-niveau-indeling. Vooral tussen de kaarten van de hogere niveaus was het verschil in moeilijkheidsgraad nogal klein. Dit weerspiegelt het kleine verschil in leesindex A tussen de hogere AVI-niveaus (zie paragraaf 2.1.2). Deze resultaten ten aanzien van de schaling van de toetskaarten roepen vragen op over de koppeling tussen (de moeilijkheidsgraad van) de leesstof en de AVI-niveaus. Dit aspect viel buiten het bestek van dit proefschrift, maar we zullen in paragraaf 5.2.3 ingaan op mogelijk onderzoek daarnaar.

De resultaten betreffende de schaling van de toetskaarten geven aan dat we niet voor het gehele continuüm van technische leesvaardigheid waarvoor de AVI-toets bedoeld is, over de meest optimale teksten beschikken. Doordat de afstanden in moeilijkheidsgraad tussen de kaarten niet gelijk zijn, is niet voor ieder deel van de onderliggende schaal de meest geschikte meting mogelijk. Voor de niet optimale volgorde en afstand in moeilijkheidsgraad is gecompenseerd via het vastzetten van de grenswaarden. Voor sommige kaarten heeft deze procedure ertoe geleid dat het verschil tussen de grenswaarden voor voldoende en goed klein is, waardoor kleine verschillen in leestijden tot een verschil van één of twee AVI-niveaus zouden kunnen leiden. Dat dit gevolgen kan hebben voor de betrouwbaarheid van de toets, spreekt voor zich.

Hoe is nu de *betrouwbaarheid* van de AVI-toetskaarten? Uit het onderzoek is gebleken dat beide versies van de AVI-toets beschouwd kunnen worden als volledig parallelle toetsen; zowel de verwachte toetsscores (de gemiddelden) als de geobserveerde score-varian-ties zijn gelijk. De betrouwbaarheid van beide versies, geschat met behulp van de Lisrel-modellen, ligt tussen 0,85 en 0,91.

Deze cijfers, gebaseerd op de *mate van samenhang* tussen beide versies, lijken niet te worden ondersteund door de gegevens over de *mate van overeenstemming* tussen beide versies op toetsniveau. Immers, in slechts ongeveer 62% van de gevallen was het AVI-niveau volgens versie A gelijk aan het niveau volgens versie B. Op kaartniveau daarentegen zijn de resultaten redelijk bevredigend te noemen. Vooral voor de vier laagste niveaus en de twee hoogste niveaus is het percentage consistente uitkomsten hoog. Dit wijst erop dat de toetskaarten het meest betrouwbaar zijn voor het meten van de technische leesvaardigheid in de periode dat kinderen deze vaardigheid verwerven voor de diverse woordtypen (AVI-niveaus 1 tot en met 4) en in de periode dat kinderen de vaardigheid dusdanig beheersen dat ze zonder problemen teksten automatisch kunnen decoderen (AVI-niveau 8 en hoger).

De vraag die we ons nu stellen is: hoe kan de overeenstemming matig zijn, terwijl de samenhang tussen beide versies hoog is en de gemiddelden op beide versies gelijk zijn? Een probleem voor de AVI-toets is uiteraard de al eerder genoemde opbouw in moeilijkheidsgraad van de toetskaarten, waardoor voor sommige toetskaarten het verschil in leestijd tussen onvoldoende, voldoende en goed erg klein is. Het feit dat we, ten behoeve van de hanteerbaarheid in de praktijk, de feitelijk gevonden grenswaarden voor de leestijden hebben afgerond tot een veelvoud van vijf seconden, zal ook bijdragen aan de inconsistenties. Daarnaast worden de inconsistente uitkomsten eveneens veroorzaakt doordat van een continue schaal een discrete schaal is gemaakt. De onderliggende continue schaal van technische leesvaardigheid is immers in negen AVI-niveaus opgedeeld. Met name voor een leerling wiens technische leesvaardigheid in de buurt van de scheiding tussen niveaus X en X+1 ligt, geldt dat de kans groot is dat hij de ene keer in niveau X wordt ingedeeld en een andere keer in niveau X+1. Dit probleem komt ook voor bij andere toetsen wanneer de behaalde scores worden omgezet in decielscores of stanines en is dus niet inherent aan de AVI-toets zelf.

De mate van overeenstemming tussen beide versies zou overigens enigszins vergroot kunnen worden door de afnameprocedure van de AVI-toets te versoepelen. In deze soepelere procedure wordt alleen gebruik gemaakt van de 'voldoende'-grenzen voor leestijden en het aantal fouten. Men toetst vervolgens door tot een leerling 'onvoldoende' behaalt op een kaart; het AVI-niveau is dan het niveau van de vorige kaart, waarop nog wel 'voldoende' behaald werd. Deze soepelere procedure verhoogde het percentage overeenstemming tussen versie A en B tot 68%. Deze toename van 6% vraagt echter wel om een grotere tijdsinvestering die alleen in twijfelgevallen de moeite waard is. Ondanks de problemen rond de consistentie van de uitkomsten, kunnen we concluderen dat de twee versies van de AVI-toets als redelijk betrouwbare paralleltoetsen beschouwd mogen worden.

Het laatste onderzochte aspect van de psychometrische kwaliteit betrof de *validiteit* van de AVI-toetskaarten: meet de toets wat deze beoogt te meten? Nagegaan is of de gevon-

den verbanden tussen de AVI-toets en andere toetsen en variabelen overeenkomen met wat op grond van de theorie verwacht kon worden. De resultaten wijzen erop dat de AVI-toets valide is. De samenhang met de Eén-Minuut-Test, een andere toets voor technisch lezen, was hoog. We wijzen erop dat de manier waarop de Eén-Minuut-Test en de AVI-toets de decodeervaardigheid meten sterk vergelijkbaar is, namelijk via het hardop lezen van woorden respectievelijk teksten. Niet is nagegaan in welke mate de decodeerbaarheid gemeten met de AVI-toets samenhangt met decodeervaardigheid die op een andere wijze is vastgesteld, bijvoorbeeld met behulp van lexicale decisiетaken. Uit onderzoek blijkt dat de hardop-leesmethoden hoog samenhangen met lexicale decisiетaken (zie paragraaf 1.5.1). Voor een klein deel van de leerlingen, die uitspraakproblemen hebben, zouden andere tests dan de Eén-Minuut-Test en de AVI-toetskaarten de voorkeur verdienen voor het bepalen van de decodeervaardigheid.

De samenhang met toetsen voor begrijpend lezen en woordenschat (die andere construaties beogen te meten) was zoals verwacht minder hoog. De samenhang tussen technisch lezen enerzijds en begrijpend lezen en woordenschat anderzijds was het hoogst aan het begin van het leesonderwijs en nam af naarmate het onderwijs vorderde, hetgeen overeenkomt met wat we op basis van de theorie mochten verwachten. In paragraaf 5.2.3 zullen we wat dieper ingaan op de theoretische consequenties van deze resultaten.

5.2.2 De waarde voor de onderwijspraktijk

Een van de belangrijkste redenen om nieuwe AVI-toetskaarten te ontwikkelen was de vraag ernaar vanuit de onderwijspraktijk. De vraag die we in deze paragraaf willen beantwoorden is: welke waarde hebben de nieuwe AVI-toetskaarten voor die onderwijspraktijk?

Uit het onderzoek blijkt dat de AVI-toets een valide en redelijk betrouwbaar instrument is ter bepaling van de technische leesvaardigheid van de leerlingen. Net als de Eén-Minuut-Test is de AVI-toets geschikt om de ontwikkeling van de technische leesvaardigheid vast te stellen. De samenhang tussen beide toetsen is hoog. Beide kunnen dan ook als indicator voor de technische leesvaardigheid worden gezien, die als een aanvulling op elkaar gehanteerd kunnen worden.

Om een goed oordeel te kunnen geven over de prestaties van een leerling mag nooit uitgegaan worden van slechts één toetsscore. Er is immers altijd sprake van een momentopname, waarbij allerlei andere factoren de prestaties van een leerling beïnvloeden. De leerling kan net toevallig een goede of slechte dag hebben, onverwachte gebeurtenissen verstoren de toetsafname, enzovoorts. Daarom is het ook belangrijk verschillende instrumenten en procedures te gebruiken om tot een valide oordeel te komen over de technische leesvaardigheid van een leerling. Aangezien de AVI-toetskaarten en woordentests als de Eén-Minuut-Test en de Drie-Minuten-Toets het technisch lezen op een iets andere manier meten, is het aan te raden zowel een woordentest als de AVI-toetskaarten af te nemen om zo meer volledig zicht te krijgen op de technische leesvaardigheid van de leerlingen.

Vanuit het onderwijsveld wordt veelvuldig de vraag gesteld welk AVI-niveau hoort bij een score op de Eén-Minuut-Test of de Drie-Minuten-Toets. De achtergrond van deze vraag is meestal dat men uitgaande van een score op de Eén-Minuut-Test of de Drie-Minuten-

Toets het AVI-niveau van de leerlingen wil bepalen, om op basis daarvan leesstof te selecteren. Het afnemen van één van deze twee woordentests kost immers minder tijd dan het afnemen van de AVI-toets. Uit ons onderzoek is gebleken dat de samenhang tussen de Eén-Minuut-Test en de AVI-toets hoog is, maar niet perfect: de correlaties liggen duidelijk lager dan 1. Het feit dat beide toetsen niet volledig uitwisselbaar zijn, betekent voor de onderwijspraktijk dat de scores op de Eén-Minuut-Test niet rechtstreeks kunnen worden omgezet in een AVI-niveau. Hetzelfde geldt overigens voor de Drie-Minuten-Toets van het Cito (Gillijns & Visser, 1995). Wel is het mogelijk om op basis van de score op een van beide woordentests te bepalen met welke AVI-toetskaart gestart kan worden teneinde het AVI-niveau te bepalen (zie Visser, Van Laarhoven & Ter Beek, 1996). Scores op de Eén-Minuut-Test en de Drie-Minuten-Toets geven, net als een score op de AVI-toets, aan dat leerling A beter decodeert dan leerling B, en dat leerling C op tijdstip X+1 beter decodeert dan op tijdstip X. De meerwaarde van de AVI-toets is gelegen in het feit dat een score tevens aangeeft dat een leerling toetskaarten van een bepaalde moeilijkheidsgraad aan kan. Alleen met deze toets kan het AVI-niveau op een betrouwbare en valide manier worden bepaald.

Daarmee komen we bij het doel waarvoor de toetskaarten in de jaren zeventig ontwikkeld zijn: het afstemmen van de leesstof op de technische leesvaardigheid van de leerlingen om zo te kunnen differentiëren in het leesonderwijs. De toetskaarten voldoen telkens aan bepaalde criteria die het AVI-niveau van een tekst bepalen. De onderzoeksresultaten betreffende de schaling van de toetskaarten geven aan dat de indeling van de toetskaarten naar AVI-niveau niet altijd een goede weergave is van de daadwerkelijk ervaren moeilijkheidsgraad van de tekst. Onderzoek naar de relatie tussen (de criteria voor) de AVI-niveaus en de moeilijkheidsgraad van leesboeken viel buiten het bestek van dit proefschrift, maar de resultaten geven wel aan dat dit belangrijk is. In paragraaf 5.2.3 zullen we daarop terug komen.

Dat met name voor de vijf hoogste niveaus geldt dat de ervaren moeilijkheidsgraad (gemeten met behulp van de leessnelheid) niet volledig overeenkomt met de verwachte moeilijkheidsgraad is niet zo vreemd, aangezien de indeling van deze niveaus uitsluitend gebaseerd is op de Leesindex A van Brouwer. Leesbaarheidsindices zijn immers slechts een indicator van de moeilijkheidsgraad van een tekst. Daarom is het altijd belangrijk om ook andere aspecten te betrekken bij het kiezen van leesstof, zoals bijvoorbeeld de inhoud, de herkenbaarheid en de opbouw van een boek, de interesse van het kind, het doel waarvoor de tekst gekozen wordt, enzovoort. Het is ook nooit de bedoeling geweest dat het AVI-niveau het enige criterium was voor selectie van leesboeken. In de AVI-niveaulijsten worden dan ook naast het AVI-niveau enkele andere kenmerken van een boek aangegeven, zoals thema en schrijfstijl.

Vanuit de onderwijspraktijk wordt over de AVI-toets vaak de opmerking gemaakt dat "vergeten wordt" om na te gaan of de leerling de tekst ook heeft begrepen. Men suggereert dan dat bij de AVI-toets begripsvragen gemaakt zouden moeten worden en dat het resultaat op deze begripsvragen mee moet worden gewogen bij het bepalen van het AVI-niveau (Faber, 1995). Deze kritiek komt wellicht voort uit de gedachte dat lezen in de praktijk niet is op te splitsen in deelvaardigheden. Men wil een test die representatief is

voor een doorsnee leessituatie waarin verschillende leesaspecten tegelijk een rol spelen. Om een aantal redenen is het echter niet wenselijk om begripsvragen toe te voegen aan de AVI-toets.

Op de eerste plaats is de AVI-toets bewust geconstrueerd om een deelvaardigheid, namelijk het technisch lezen, zo betrouwbaar en valide mogelijk te meten. Daarop is ook de afnameprocedure afgestemd. Het meten van begrijpend lezen stelt andere, eigen eisen aan de toetsconstructie en toetsafname.

Op de tweede plaats is het niet wenselijk om op basis van één toets een uitspraak te willen doen over de totale leesvaardigheid van een kind, ook al is dit vanuit het oogpunt van tijdsbesparing wel begrijpelijk. Om een goed oordeel te kunnen geven over de leesvaardigheid van een leerling dienen verschillende toetsen gebruikt te worden (voor begrijpend lezen, technisch lezen en woordenschat) ter aanvulling van de observaties van de leerkracht die betrekking kunnen hebben op het totaal. Objectieve tests kunnen probleemgebieden aanwijzen waarop extra instructie nodig is.

Wellicht is de diepere reden voor de vraag naar begripsvragen gelegen in het idee dat door de AVI-toets te beperken tot het technisch lezen, voorbij wordt gegaan aan het feit dat het uiteindelijke doel van lezen tekstbegrip is. Maar het maken van een toets voor technisch lezen betekent niet dat we lezen gelijk stellen aan technisch lezen. Het betekent wel dat we vinden dat technisch lezen een van begrijpend lezen en woordenschat te onderscheiden vaardigheid is die afzonderlijke aandacht verdient. Uit het onderzoek is duidelijk gebleken dat de technische leesvaardigheid een belangrijke invloed heeft op het leesbegrip. En vanuit deze visie is het juist om, naast toetsen voor begrijpend lezen en woordenschat, ook een toets voor technisch lezen te construeren en af te nemen.

Leerkrachten kunnen de toets hanteren om problemen met technisch lezen te signaleren en gericht extra instructie aan te bieden. Hierop komen we nog terug.

De validiteit en betrouwbaarheid van de AVI-toets, zoals die in dit onderzoek is vastgesteld, gelden altijd onder de voorwaarde dat men zich houdt aan de toetsinstructie. Ook de criteria voor de niveaubepaling dienen nauwkeurig gehanteerd te worden. Dit laatste was bij de oude AVI-toetskaarten niet altijd het geval, waardoor leerlingen met eenzelfde prestatie op een toetskaart een verschillend AVI-niveau toegewezen kregen. Indien bij de nieuwe AVI-toetskaarten wederom andere criteria in omloop gebracht worden, zijn we feitelijk weer terug bij af en komt opnieuw de validiteit en betrouwbaarheid van de indeling naar AVI-niveau in het geding.

Zoals we aan het begin van deze paragraaf hebben aangegeven, is de AVI-toets herzien mede vanuit de vraag vanuit de onderwijspraktijk. Op welke aspecten is de nieuwe AVI-toets nu een verbetering van de oude?

Op de eerste plaats zijn voor de nieuwe AVI-toetskaarten landelijke normeringsgegevens beschikbaar. Door de scores van de leerlingen te vergelijken met deze normgegevens kan een leerkracht een indicatie krijgen van de technische leesvaardigheid van de leerlingen. Gesignaleerd kan worden welke leerlingen tot de groep risico-leerlingen behoren en dus extra aandacht en instructie behoeven om te voorkomen dat zij een te grote achterstand oplopen. De beschikbaarheid van deze gegevens is een duidelijke verbetering ten opzichte van de oude AVI-toetskaarten. Voor de oude AVI-toetskaarten waren dergelijke gegevens niet beschikbaar. Wel werden in de praktijk verschillende normen gehanteerd, bij-

voorbeeld dat aan het eind van groep 3 AVI-3 gehaald moest worden. De oorspronkelijke indeling van de negen AVI-niveaus door Van den Berg en Te Lintelo (1977b; zie ook hoofdstuk 1 en 2 van dit proefschrift)) bestempelde de eerste vier niveaus voor het aanvankelijk lezen (groep 3 en een deel van groep 4) en de overige vijf niveaus voor het voortgezet technisch lezen (groep 4 tot en met groep 6). Uit de normeringsgegevens van de nieuwe AVI-toetskaarten kan afgeleid worden dat AVI-niveau 4 door de gemiddelde leerling nog niet beheerst wordt aan het einde van het aanvankelijk leesonderwijs.

De nieuwe toets is ook op een ander punt verbeterd ten opzichte van de oude toetskaarten. In een onderzoek naar de oude toetskaarten bleek dat relatief weinig leerlingen op niveau 6 en 8 lazen en relatief veel leerlingen op niveau 7; de frequentieverdeling was tweetoppig. De verdeling van de AVI-niveaus bepaald met de nieuwe toetskaarten is één-toppig en verloopt geleidelijk in de loop van de basisschoolperiode.

Tot slot zijn de nieuwe AVI-toetskaarten samen met de Eén-Minuut-Test opgenomen in het Nijmeegse Leerlingvolgsysteem LISKAL (Aarnoutse, Van Leeuwe, Oud, Voeten & Van Kan, 1997) als indicatoren voor de technische leesvaardigheid van de leerlingen. In dit leerlingvolgsysteem zijn tevens toetsen opgenomen voor de domeinen Begrijpend Lezen, Woordenschat, Spelling en Rekenen. Het leerlingvolgsysteem biedt de mogelijkheid om zowel de relatieve ontwikkeling als de absolute ontwikkeling van de prestaties van de leerlingen weer te geven. De relatieve ontwikkeling betreft de vergelijking van de prestaties van een leerling met de gemiddelde ontwikkeling in de landelijke steekproef. De absolute ontwikkeling betreft de vergelijking van de prestaties van een leerling met de eigen, vroegere ontwikkeling. Daarbij wordt tevens een voorspelling gedaan over de te verwachten ontwikkeling in de nabije toekomst. Op grond van longitudinale gegevens kan met behulp van scores op de AVI-toetskaarten en de Eén-Minuut-Test worden voorspeld hoe de technische leesvaardigheid van een kind zich zal ontwikkelen in de volgende periode(n). Dit biedt leerkrachten de mogelijkheid bij een voorspelde achterblijvende ontwikkeling, op tijd in te grijpen om zo te proberen de voorspelling te niet te doen.

Al het voorgaande overziend kunnen we vaststellen dat de nieuwe AVI-toetskaarten een duidelijke verbetering zijn ten opzichte van de oude: de teksten zijn aangepast aan het huidige tijdsbeeld, de validiteit en betrouwbaarheid van de toets zijn bepaald en acceptabel bevonden, er zijn landelijke normeringsgegevens beschikbaar en de nieuwe toets vormt een onderdeel van het leerlingvolgsysteem LISKAL. Tevens is met dit proefschrift een verantwoording gegeven van de constructie van de kaarten en de bijbehorende indelingscriteria voor fouten en leestijd.

5.2.3 Bijdrage aan theorievorming en suggesties voor nader onderzoek

Wanneer we de onderzoeksresultaten spiegelen aan wat al bekend is over de technische leesvaardigheid, de ontwikkeling daarvan en de relatie ervan met begrijpend lezen en woordenschat dan zien we veelal een bevestiging van bestaande inzichten. Zo lijkt inderdaad de snelheid van woordherkenning de belangrijkste indicator te zijn voor de mate waarin de decodeervaardigheid geautomatiseerd is en is de ontwikkeling daarvan het snelst aan het begin van het leesonderwijs. Op enkele resultaten willen we wat nader ingaan, namelijk de gegevens betreffende de samenhang tussen de AVI-toetskaarten en de Eén-Minuut-Test en de samenhang tussen technisch lezen en begrijpend lezen.

In het onderzoek naar de validiteit van de AVI-toetskaarten is de samenhang onderzocht tussen deze toets en de Eén-Minuut-Test. De samenhang tussen beide toetsen is erg hoog (0,81-0,90), maar tot en met groep 6 (na drie jaar leesonderwijs) zijn de voor onbetrouwbaarheid gecorrigeerde correlaties niet gelijk aan één en zijn de toetsen dus niet volledig uitwisselbaar. Het belangrijkste verschil tussen beide toetsen is het gebruikte testmateriaal: gehele teksten versus een rij losse woorden. Het hardop lezen van teksten vereist blijkbaar net iets anders dan het hardop lezen van losse woorden. Het is zeer plausibel dat in het eerste geval de context toch op de een of andere manier een rol speelt. In hoofdstuk 1 hebben we twee mogelijke contexteffecten besproken, namelijk een bewust contexteffect en een automatisch contexteffect. Beide effecten kunnen een verklaring zijn voor het niet volledig overeenkomen van prestaties op een test waarin woorden worden gebruikt en een test die bestaat uit samenhangende teksten. Bij de Eén-Minuut-Test kan de context geen rol spelen, omdat die ontbreekt. Bij de AVI-toets kunnen beide effecten optreden, aangezien er wel degelijk een context aanwezig is. Het eerste contexteffect, waarbij woordherkenning wordt gefaciliteerd door verwachtingen van de lezer, treedt vooral op bij beginnende lezers. Hoe meer de decodeervaardigheid geautomatiseerd wordt, des te minder maken lezers (bewust) gebruik van de context bij het herkennen van woorden. Als dit contexteffect verantwoordelijk is voor het onderscheid tussen de AVI-toets en de Eén-Minuut-Test, dan zouden we mogen verwachten dat de correlatie tussen deze toetsen toeneemt in de tijd. Dit bleek echter niet het geval te zijn; de correlaties waren in groep 6 lager dan in groep 4 en 5. Een eerste verklaring voor de lagere correlaties in groep 6 is het plafondeffect dat met name in groep 6 optreedt bij de AVI-toetskaarten, waardoor de correlaties eveneens afnemen. Bij het lezen van teksten kan ook het tweede contexteffect optreden, waarbij door de herkenning van een woord 'automatisch' andere nabijgelegen representaties enigszins geactiveerd worden. Dit effect treedt juist op bij meer ervaren lezers. De aanwezigheid daarvan is een tweede mogelijke verklaring voor het feit dat ook aan het eind van groep 6 nog steeds een duidelijk onderscheid bestaat tussen de AVI-toets en de Eén-Minuut-Test.

Op basis van het onderzoek kunnen we geen antwoord geven op de vraag of bij het decoderen van losse woorden daadwerkelijk andere processen een rol spelen dan bij het decoderen van samenhangende teksten. Het zou interessant zijn om na te gaan of het onderscheid tussen het decoderen van losse woorden en decoderen van teksten ook bij volwassen lezers nog aanwezig is. Daarbij zou ook nagegaan moeten worden in hoeverre deze decodeervaardigheid bij teksten meer samenhangt met begrijpend lezen dan de decodeervaardigheid bij losse woorden. Indien dat inderdaad het geval is, dan zou dit een aanwijzing kunnen zijn dat lezers die goed 'automatisch' gebruik kunnen maken van de context in het voordeel zijn bij het snel lezen van teksten.

Daarmee komen we op de samenhang tussen de technische leesvaardigheid en begrijpend lezen. Uit het onderzoek is gebleken dat aan het begin van het leesonderwijs de ontwikkeling van de technische leesvaardigheid een positieve invloed heeft op het begrijpen van een tekst. Goede en zwakke begrijpend lezers verschillen met name van elkaar in de snelheid van woordherkenning. Het onderzoek bevestigt dat het verband tussen technisch lezen en begrijpend lezen samenhangt met de leeservaring en de leesvaardigheid:

de correlaties zijn het hoogst in de lagere leerjaren. Al vanaf het najaar van groep 5 lijkt het erop dat een verdere toename in leessnelheid geen invloed meer heeft op de begrip-pende leesvaardigheid, of beter gezegd op de rangorde in begrip-pend lezen. Degenen die op dat moment goed zijn in begrip-pend lezen blijven dat ook, de zwakkeren blijven rela-tief de zwakkeren. Wel is het zo dat de relatie tussen technisch lezen en begrip-pend le-zen hoog blijft. De technische leesvaardigheid aan het begin van het leesonderwijs is dan ook een goede voorspeller van de uiteindelijke vaardigheid in begrip-pend lezen. Dit betekent dat aan het begin van het leesonderwijs veel aandacht besteed moet worden aan de technische leesvaardigheid. Tijdige signalering van uitvallers is daarbij belangrijk. Immers de bevinding dat vanaf groep 5 een toename in de technische leesvaardigheid geen effect meer heeft op de begrip-pende leesvaardigheid geeft aan dat het na groep 4 te laat is: het kwaad is dan al geschied.

Naast de hierboven beschreven samenhang tussen technisch lezen en begrip-pend lezen, bevestigen de onderzoeksresultaten dat technische leesvaardigheid duidelijk te onder-scheiden is van de vaardigheid om met begrip teksten te lezen. Toetsen voor technisch lezen meten een andere vaardigheid dan toetsen voor begrip-pend lezen en woordenschat. De toetsen voor technisch lezen verschillen echter ook 'meet-technisch' gezien van de toetsen voor begrip-pend lezen en woordenschat: bij het eerste type toetsen speelt de factor tijd of snelheid een grote rol. Voor de toetsen voor begrip-pend lezen en woorden-schat krijgen leerlingen meestal voldoende tijd om de toets af te maken, terwijl bij de toetsen voor technisch lezen de snelheid en dus de tijd een allesbepalende factor is. Het zou interessant zijn na te gaan wat de samenhang is tussen snelheid van decoderen en snelheid van begrip. In onze maatschappij, waar het snel opbouwen van kennis een gro-te rol speelt, zou het snel begrijpen van teksten weleens erg belangrijk kunnen zijn. Leerlingen en studenten dienen vaak grote hoeveelheden studieteksten te verwerken binnen een beperkte tijdsperiode. Onderzoek naar de mate waarin de snelheid van decode-ren daarbij van belang is, zou zeer zinvol zijn. Een dergelijk onderzoek zou bijvoorbeeld gehouden kunnen worden onder middelbare scholieren of bij studenten.

Specifiek gericht op de AVI-toetskaarten zijn er een aantal aspecten die nader onder-zocht kunnen worden.

Op de eerste plaats zou het gewenst zijn onderzoek te doen naar de relatie tussen de moeilijkheidsgraad van de leesstof en de AVI-niveaus. In het huidige onderzoek stond de relatie tussen de toetskaarten en de technische leesvaardigheid van de leerlingen cen-traal. Zowel de toetskaarten als de prestaties van de leerlingen zijn afgebeeld op een onderliggend continuüm van technische leesvaardigheid. De volgende stap is om ook leesstof een plek te geven op dit continuüm, zodat duidelijk wordt welke technische leesvaardigheid een leerling nodig heeft om de betreffende tekst 'technisch' te kunnen lezen. Dit gebeurt nu via het toekennen van de AVI-niveaus aan de leerlingen met be-hulp van de toetskaarten en aan de leesstof met behulp van de AVI-indelingscriteria voor teksten. In het voorgestelde onderzoek zou in feite een validering van deze praktijk plaatsvinden, waarbij het uiteraard mogelijk is dat de huidige indeling van teksten naar AVI-niveaus niet geheel juist is.

Staphorsius en Verhelst (1997) hebben een onderzoek uitgevoerd naar de vereiste lees-techniek voor teksten. Dit onderzoek geeft belangrijke aanknopingspunten voor het door ons voorgestelde onderzoek. Feitelijk kunnen we het onderzoek naar de validering van de AVI-niveaus van teksten zien als een uitbreiding van het onderzoek naar de schaling van de AVI-toetskaarten. Het zou mogelijk zijn om hetgeen we nu voorstellen uit te voeren op de beschikbare gegevens van de toetskaarten, ware het niet dat het aantal teksten (18) dan erg klein is. In het voorgestelde onderzoek zou een representatieve steekproef leerlingen een ruimere hoeveelheid teksten hardop dienen te lezen, waarbij voor iedere tekst de leestijd wordt vastgelegd. Bij deze leerlingen wordt tevens de AVI-toets afgenomen. Vervolgens kunnen de teksten, inclusief de toetskaarten, onderling worden geschaald, gebruikmakend van hetzelfde aangepaste Rasch-model als voor de schaling van de toetskaarten in dit proefschrift is toegepast. Van iedere tekst uit de steekproef wordt zo duidelijk welke positie in moeilijkheidsgraad deze heeft ten opzichte van de verschillende AVI-toetskaarten.

Vervolgens kan nagegaan worden welke tekstenkenmerken belangrijk zijn voor het bepalen van de relatieve moeilijkheidsgraad van de teksten, waarmee de indeling van teksten naar AVI-niveaus getoetst kan worden. De eerste vier niveaus van de AVI-toets zijn duidelijk gebaseerd op woordkenmerken, de overige niveaus op Leesindex A van Brouwer. In paragraaf 2.1.2 hebben we de vraag gesteld of het wel correct is om een leesbaarheidsformule, die bedoeld is voor het vaststellen van de begrijpelijkheid van teksten, te gebruiken voor het bepalen van de technische moeilijkheidsgraad. Daarbij hebben we wel opgemerkt dat in dergelijke formules, en ook in Leesindex A, technische aspecten van een tekst zijn gebruikt, namelijk woord- en zinslengte. In ieder geval is van de woordlengte bekend dat deze samenhangt met de snelheid van woordherkenning. Het is de vraag of dat ook geldt voor de zinslengte.

Staphorsius en Verhelst (1997) hebben een index ontwikkeld specifiek voor de leesteknik, de CILT. In deze formule zijn de aspecten woordlengte en woordfrequentie de variabelen die de moeilijkheidsgraad voor technisch lezen bepalen. Vergeleken met de door Staphorsius (1994) ontwikkelde leesindex voor de leesbaarheid CLIB (zie paragraaf 2.1.1) ontbreken de variabelen zinslengte en diversiteit van de woordenschat. Dit zou erop wijzen dat de zinslengte niet belangrijk is voor de moeilijkheidsgraad voor technisch lezen. In het door ons voorgestelde onderzoek zouden we desondanks de variabelen zinslengte en diversiteit van de woordenschat, naast woordlengte en woordfrequentie, willen betrekken. Gezien de opbouw van de eerste vier AVI-niveaus zouden ook variabelen meegenomen moeten worden die het aandeel één-, twee-, drie- en meerlettergrepen weergeven.

Analoog aan het onderzoek van Staphorsius en Verhelst zullen vervolgens regressie-analyses uitgevoerd dienen te worden, met bovengenoemde tekstenkenmerken als onafhankelijke variabelen en de moeilijkheidsgraad van de teksten zoals die resulteren uit het aangepaste Rasch-model als afhankelijke variabele. Op deze manier kan nagegaan worden in hoeverre de op deze manier gevonden index overeenkomt met de huidige AVI-niveau-indeling van teksten.

Toepassing van het model zal tevens een schaling opleveren van de leerlingen naar technische leesvaardigheid. Vervolgens dient vastgesteld te worden voor de leerlingen van welk AVI-niveau een tekst geschikt is. Hiervoor zijn diverse gegevens beschikbaar: de

positie van de tekst ten opzichte van de toetskaarten en de relatie tussen de AVI-niveaus van de leerlingen respectievelijk de positie van de leerlingen op de onderliggende schaal voor technische leesvaardigheid en de benodigde leestijd voor de tekst. De hamvraag blijft echter hoe uiteindelijk vastgesteld kan worden wanneer een tekst geschikt is voor een bepaald AVI-niveau. Staphorsius en Verhelst hebben net als Staphorsius (1994) uiteindelijk gebruik gemaakt van leerkrachtoordelen om te bepalen voor welke groepen leerlingen de teksten geschikt zijn. Dit is uiteraard een subjectief criterium, maar wellicht op dit moment het meest hanteerbare om tevens toe te passen in het door ons voorgestelde onderzoek.

Een tweede aspect dat onderzocht kan worden is het gebruik van de AVI-toets zelf. Uit gesprekken met leerkrachten en onderwijsbegeleiders blijkt dat de toets nogal eens als een informele test wordt gebruikt. Naast het bepalen van het AVI-niveau van de leerling, leveren volgens hen de observaties tijdens de toetsafname waardevolle informatie op over het type fouten dat gemaakt wordt en de manier waarop een leerling teksten leest. Een onderzoek hiernaar zou zinvolle informatie kunnen opleveren voor de onderwijspraktijk.

Kijkend naar het doel van de AVI-niveaus, het selecteren van geschikte leesstof die aansluit bij de technische leesvaardigheid van leerlingen, zou onderzocht kunnen worden hoe bruikbaar de AVI-niveaus in de praktijk zijn voor het kiezen van boekjes. Te beantwoorden vragen daarbij zijn: hoe worden de AVI-niveaus gebruikt in de praktijk bij het selecteren van boeken, kunnen leerlingen boeken aan die ver boven hun AVI-niveau liggen en wanneer geldt dat, welke andere criteria spelen een rol, enzovoort. Dit onderzoek zou met name voor de onderwijspraktijk relevante gegevens op kunnen leveren: hoe strikt moet men zich houden aan de AVI-niveaus bij de selectie van boeken, welke andere criteria spelen daarbij een rol, et cetera. Hierbij aansluitend zou tevens onderzocht kunnen worden op welke wijze de AVI-niveaus in de lessituatie worden gebruikt. Een punt van aandacht daarbij is het zogenaamde niveaulezen: op welke manier wordt dit op school gerealiseerd, welke effecten heeft het op de leesvorderingen van leerlingen, enzovoort.

Uit het onderzoek naar de nieuwe AVI-toetskaarten is gebleken dat deze een redelijk betrouwbaar en valide instrument vormen om de technische leesvaardigheid van de leerlingen te meten. Dit laatste, het meten van de decodeervaardigheid van de leerlingen, is zeker in de eerste jaren van het basisonderwijs erg belangrijk. Tijdige signalering van zwakke technische lezers is noodzakelijk, aangezien uit het onderzoek blijkt dat vanaf groep 5 een toename in technische leesvaardigheid nauwelijks effect meer heeft op de prestaties bij begrijpend lezen. De AVI-toets is een goed instrument om zwakke decodeerders tijdig te signaleren, zodat de leerkracht deze leerlingen extra instructie kan aanbieden. Een investering in de technische leesvaardigheid in de eerste jaren van het leesonderwijs betekent een blijvend voordeel voor de leerlingen gedurende de rest van hun lees-loopbaan.

SUMMARY

AVI reading level:

The scaling, norming, and validation of the new AVI test cards

The AVI test cards

The AVI test cards originally constituted part of a project undertaken by the KPC in 1972 to develop procedures and resources for the individualization of elementary reading instruction (Analyse Van Individualiseringsvormen, AVI). This included individualization of the reading technique, reading comprehension and the reading experience. Nine special AVI levels were created for decoding instruction. These levels were used to categorize reading books according to the degree of difficulty and also to categorize the decoding skill of students. The AVI cards were developed particularly for the latter purpose. The cards can be used to determine the word-recognition skills of children and to select reading materials corresponding to this level for decoding instruction. In this dissertation, the reliability and validity of the new AVI test cards for the determination of the decoding skill of students were examined.

The AVI test consists of one card for each AVI level or a total of nine cards. Each AVI level represents a particular degree of technical difficulty. Each card contains a text presenting a complete story. Two different versions of the AVI test, the A and the B versions, are available.

The administration of the test occurs on an individual basis. The student must read the text presented on the card out loud. The number of errors and the time (in seconds) that the student needs to read the text specify the level of reading achievement or the student's AVI level. For each card, specific criteria for the number of errors and the reading time determine whether the level of reading was unsatisfactory, satisfactory, or good. In the case of *satisfactory* achievement, further administration of the test is not necessary, and the student is assigned to the relevant AVI level. In the case of *unsatisfactory* achievement, the student is administered the card from one level lower; in the case of *good* achievement, the student is administered the card from one level higher. In such a manner, administration of the test continues until a *satisfactory* level of performance is achieved.

The AVI test is intended to determine the degree to which the student's decoding skill has been automatized. By automatized, we mean that a word is accurately, automatically, and -- in particular -- quickly recognized. The speed of word recognition is, in this context, the best indicator of the degree automatization for the skill. With the aid of the AVI test, both the accuracy (number of errors) and the speed (reading time) of the student's decoding skill are determined. Maximum decoding speed is not attained for all words at the same point in development: short, regularly spelled, and frequently occurring words are read more quickly and automatically than long, irregularly spelled, and less frequently occurring words. The AVI test is thus composed of texts with increasing degrees of difficulty along these lines. Levels 1 through 4 vary with respect to the degree of difficulty from one-syllable words to three-or-more-syllable words. Levels 5

through 9 vary with respect to the Reading Ease Formula A from Brouwer (1963) which has word and sentence length as the main determinants of difficulty.

Revision of the AVI test cards

In the period 1992-1994, new AVI test cards were developed. The initial reason for this revision was the fact that the text and illustrations -- developed in the 1970s -- were no longer attuned to the experiences of the children taking the test. In addition, the criteria used to establish the different degrees of difficulty for the first four AVI levels were almost exclusively based on the old version of the basic reading program 'Veilig Leren Lezen' (Learning to Read Safely). A third reason for revising the AVI test cards was that the existing criteria for the number of errors and reading time did not have sufficient empirical justification. Many schools appeared to use criteria other than the ones accompanying the test, moreover. In such a manner, students with exactly the same performance on a particular test card might be assigned to different AVI levels depending on the criteria being used.

Just as for the old AVI test, it was decided to use written text as the test material in the new AVI test. The first reason for this decision was that the test in this form has received broad acceptance in the elementary schools. A second reason concerns the ecological validity of the test: By using written text, the test situation clearly resembles the actual reading situation in which one also reads cohesive text. In addition, the AVI test with written text can supplement such word-based tests for decoding as the One Minute Test and Three Minute Test (*Eén-Minuut-Test, Drie-Minuten-Toets*). A final reason for using entire texts as test material was the fact that beginning readers are known to use the reading context for decoding. To the degree that the decoding skill becomes more automatic, the influence of the context declines.

The construction of the new AVI test cards occurred in a number of phases. In the first phase, the basic-reading-method-dependent criteria were removed from the first four AVI levels; the criteria for the other five levels remained the same. New texts were then written for each of the nine AVI levels, and it was attempted to gear these texts as much as possible to the experiences of children today. The texts were next presented to a number of reading experts for evaluation. The commentary from these experts was then processed by a copy editor while clearly keeping the criteria for the different AVI levels in mind. The resulting texts were then piloted at a school and evaluated for their degree of difficulty by 25 teachers. On the basis of this initial piloting, the texts were again adapted and thereafter tested at seven elementary schools. After some final revision, the scaling, norming, and validation of the test cards were undertaken as part of a longitudinal study of school progress within the Department of Educational Sciences at the University of Nijmegen.

Research design

Research questions

The following questions stood central to the research reported in this dissertation.

- 1 Do the new AVI test cards increase in degree of difficulty?
- 2 What is the reliability of the AVI test cards?

- 3 What is the distribution of the AVI levels across the elementary school period?
- 4 Can the A and B versions of the AVI test be considered parallel?
- 5 What is the relation between the AVI test and the One Minute Test?
- 6 What is the relation between the AVI test and tests of reading comprehension and vocabulary?

Data collection

For scaling, norming, and validation, the AVI test cards were administered to a sample of students from elementary grades 1 through 4. The sample was part of a longitudinal study at the Department of Educational Sciences in Nijmegen. In this study, three student cohorts starting grades 1, 2, and 3 in September 1991 were followed with 800-1000 students per cohort. It should be noted that urban schools with a significant foreign population and otherwise disadvantaged Dutch students were under-represented in this sample. For the evaluation of the AVI test cards, the data collected in the spring and fall of 1992 and the spring of 1993 were used.

In March 1992 and March 1993, both the A and B versions of the new AVI test were administered. In October 1992, only the A version was administered. In principle, the students received three consecutive test cards close to their presumed level of reading; the test cards were always from the same version of the AVI test. The reading time and number of errors per test card were used to scale and norm the test cards.

In addition to the AVI levels, the scores on another test of decoding -- the One Minute Test -- and various tests of reading comprehension and vocabulary were used to validate the test cards. Missing data were replaced using of the imputation procedure from Prelis 2 (Jöreskog & Sörbom, 1993b). This occurred per cohort per testing period provided that a minimum of four test scores were available for a particular student in the relevant period (with a total of 6 scores per testing period possible). Students with too many missing scores were excluded from the analyses. Closer examination of the results showed that the data were not missing at random. For particularly the students in cohort 1, those included in the validation analyses appeared to perform better on the tests of decoding than those not included.

Research results

Determination of boundary values

The first step was to establish the boundary values for the number of errors and reading times for a particular AVI level. Three levels of performance were distinguished per card: unsatisfactory, satisfactory, and good. For both the number of errors and reading times, thus, two boundary values for the transition from unsatisfactory to satisfactory performance and satisfactory to good performance had to be established.

The test cards were initially scaled according to an extension of the Rasch-model (Rasch, 1980) for analyzing reading-speed data (Jansen, 1997). A model incorporating both reading times and number of errors was not available. Given that reading speed is the best indicator of the degree to which decoding skill has been automatized, moreover, it was logical to take reading time as the starting point in determining the boundary values for the different AVI levels.

After the AVI test cards were scaled according to reading time, they were divided into nine AVI levels of decoding ability. The boundary values for the reading times were chosen to ensure consistency of the outcomes from the different AVI test cards: Children achieving level 2, for example, should also achieve level 1. To facilitate actual practice, the reading times were rounded to the nearest multiple of five seconds. Finally, the boundary values for the reading errors were determined in keeping with the principle that decoding speed should not occur at the cost of accuracy.

The degree of difficulty for the AVI test cards (research question 1)

The AVI test cards are based on the nine AVI levels intended to reflect both the increasing technical difficulty of the texts and the increasing decoding skills of students. Application of the adapted Rasch model enabled us to estimate the degree of difficulty for each text. This scaling showed the ordering of the test cards according to the degree of reading-time difficulty to not completely correspond to the ordering based on the AVI levels. The differences in degree of difficulty for the cards at the higher levels was particularly small. The boundary values were adjusted to compensate for the less than optimal order and small differences observed in the degree of difficulty in the AVI test cards. The scaling results for the test cards raise questions about the connection between the (difficulty of the) reading material and the AVI levels. Further research into this matter should therefore be welcomed.

The reliability of the AVI test cards (research questions 2 and 4)

The results of the present research showed that the two versions of the AVI test can be considered completely parallel; both the expected test scores (i.e., averages) and the observed score-variances were found to be equivalent for the two versions of the test. Estimated using Lisrel models, the reliability between the two versions was found to range between 0.85 and 0.91.

This range of reliability, which is based on the *degree of correlation* between the scores on the two versions of the AVI test, is not supported by the *degree of correspondence* between the two versions with regard to classification. In only about 62% of the cases was the AVI level according to version A of the test found to be equivalent to version B. At the card level, in contrast, the results can be considered fairly satisfactory. The percentage of results found to be consistent was particularly high for the four lowest levels and the two highest levels. This suggests that the test cards are most reliable for the measurement of decoding skill in the period when children are acquiring this skill for different word types (AVI levels 1 through 4) and in the period when children have mastered this skill to such a degree that they can automatically decode texts without much problem (AVI levels 8 and 9). The first source of inconsistency may be the earlier mentioned problems with the order and small differences in the degree of difficulty of the test cards, resulting in relatively little differences between the unsatisfactory, satisfactory, and good reading times. The fact that we rounded the observed boundary values in order to facilitate use of the tests in actual practice may also have contributed to the lack of consistency between the two versions of the test. Finally, the inconsistent outcomes may also be caused by the transformation of a continuous scale into a discrete scale. This problem has also been encountered when the attained scores on other tests

are converted into decile scores or stanines and is thus not inherent to the AVI test itself.

The distribution of the AVI levels across the elementary school period (research question 3)

The application of the adapted Rasch model also produced estimates of the students' decoding skill. Between March of grade 1 and March of grade 4, a clear increase in decoding skill could be observed. The growth in decoding skill was greatest at the start of reading instruction and declined in the higher grades. Comparable development could also be seen in the AVI levels. Over the course of the elementary school period, the AVI levels of the students clearly increased. In March of grade 1, 44% of the students had not as yet mastered AVI level 1. In March of grade 4, almost 50% of the students had mastered level 9 of the AVI test. The transitions from the different AVI levels were also found to occur quite regularly without large or abrupt shifts. Comparison of the average AVI level across time again showed the growth at the beginning of reading instruction to be strongest: an average level of 1.1 for students in March of grade 1 to an average level of 4.5 in March of grade 2, which represents a growth of more than three levels in one year. The average growth observed in the later stages of elementary education was weaker than that observed in the earlier stages: from an average level of 6.9 for students in March of grade 3 to an average level of 8.3 in March of grade 4, which represents a growth of somewhat more than one level in one year.

The validity of the AVI test cards (research questions 5 and 6)

Whether or not the relations between the AVI test and other tests corresponded to what might be expected on the grounds of current theory was also examined. The results showed the AVI test to be valid. The correspondence to another test of decoding, the One Minute Test, was high. The correlations (corrected for unreliability) between the latent variables from the AVI test and the One Minute Test (both measured with the aid of the A and B versions of the relevant tests) were found to lie between 0.83 and 0.90 for grades 1 through 3 and between 0.81 and 0.83 for grade 4. The correlations were reasonably stable for grades 2 and 3 and dropped slightly for grade 4. A possible ceiling effect associated with the AVI test cards could account for this lower correlation with the One Minute Test, which includes additional levels of performance. The observed correlations were not perfect (less than 0.95) which means that the tests do not measure exactly the same phenomena. By using text instead of individual words, that is, the AVI test appears to measure a slightly different aspect of decoding than the One Minute Test.

The fact that the two tests are not completely interchangeable means that the scores on the One Minute Test should not be directly translated into AVI levels for educational practice. It may certainly be possible to determine the level of AVI test card to start with on the basis of the One Minute Test.

The relations between the AVI test cards and the various tests for reading comprehension and vocabulary were first examined per cohort per testing period. The correspondence to the tests for reading comprehension and vocabulary (which are intended to measure other constructs) was, as expected, less than the correspondence to the One Minute Test. The AVI test cards clearly measure something other than the tests

for reading comprehension and vocabulary. The correlation between, on the one hand, decoding and, on the other hand, reading comprehension and vocabulary was found to be highest at the beginning of reading instruction (between 0.60 and 0.70 for grade 2) and to decrease with the progression of the instruction (between 0.55 and 0.60 for grade 3; between 0.46 and 0.55 for grade 4). This is in keeping with what we might expect on the basis of theory.

A longitudinal model of the relation of the AVI-levels to reading comprehension and vocabulary was next developed for each cohort. The results of this modelling showed the initial gains in decoding skill to positively affect early reading comprehension and vocabulary. The relation between decoding and reading comprehension is strongly associated with the degree of reading experience and reading skill: The correlations are highest in the earlier years of reading. Starting in the fall for grade 3, further increases in decoding speed no longer appear to influence reading comprehension or, more specifically, the rank ordering for reading comprehension. Those found to be good in reading comprehension at that moment remain so, and those found to be weaker at that moment also remain so. It is nevertheless the case that the relation between decoding and reading comprehension continues to be high. Decoding skill at the beginning of reading instruction is thus a good predictor of final reading-comprehension skill. This means that a lot of attention should be paid to decoding during the beginning stages of reading instruction. The early detection of those already lagging behind in this skill also appears to be important in light of the finding that increases in decoding skill starting in grade 3 no longer affect reading comprehension. The AVI test is a good instrument for the timely detection of weak decoding skill and thereby enables the provision of extra instruction before it is too late. Put differently: An investment in decoding during the beginning stages of reading instruction can continually benefit students throughout their reading careers.

6 LITERATUUR

- Aarnoutse, C.A.J. (1987). *Synoniementest, Tegenstellingentest, Voegwoordentest, Hoofdge-dachtetest. Tests voor begrijpend lezen bestemd voor groep 7 en 8 van het basisonder-wijs*. Nijmegen: Berkhout Nijmegen BV.
- Aarnoutse, C.A.J. (1988a). *Begrijpend leestest en leesattitudeschaal bestemd voor groep 4 van het basisonderwijs. Handleiding en testmateriaal*. Nijmegen: Berkhout Nijmegen BV.
- Aarnoutse, C.A.J. (1988b). *Begrippentest (woorden) en woordenschattest bestemd voor groep 3 van het basisonderwijs. Handleiding en testmateriaal*. Nijmegen: Berkhout Nijmegen BV.
- Aarnoutse, C.A.J. (1989a). *Begrijpend Leestest en leesattitudeschaal bestemd voor groep 5 van het basisonderwijs. Handleiding en testmateriaal*. Nijmegen: Berkhout Nijmegen BV.
- Aarnoutse, C.A.J. (1989b). *Begrippentest en woordenschattest bestemd voor groep 4 van het basisonderwijs. Handleiding en testmateriaal*. Nijmegen: Berkhout Nijmegen BV.
- Aarnoutse, C.A.J. (1989c). *Begrippentest en woordenschattest bestemd voor groep 5 van het basisonderwijs. Handleiding en testmateriaal*. Nijmegen: Berkhout Nijmegen BV.
- Aarnoutse, C.A.J. (1990a). *Begrijpend Leestest bestemd voor groep 6 van het basisonder-wijs. Handleiding en testmateriaal*. Nijmegen: Berkhout Nijmegen BV.
- Aarnoutse, C.A.J. (1990b). *Woordenschattest en leesattitudetest bestemd voor groep 6 van het basisonderwijs. Handleiding en testmateriaal*. Nijmegen: Berkhout Nijmegen BV.
- Aarnoutse, C.A.J. (1991). Onderwijs in lezen. In: A.J.W.M. Thomassen, L.G.M. Noordman & P.A.T.M. Eling. *Lezen en begrijpen. De psychologie van het leerproces. 2e herziene druk* (pp 223-251). Amsterdam/Lisse: Swets & Zeitlinger.
- Aarnoutse, C.A.J. (1992). Voortgezet lezen. In: Nijmeegse Werkgroep Taaldidactiek. *Taal-didactiek aan de basis. Vierde geheel herziene druk* (pp 275-315). Groningen: Wolters-Noordhoff.
- Aarnoutse, C.A.J. & Leeuwe, J.F.J. van (1986). Een verkennend onderzoek naar factoren die een rol spelen bij de verklaring van verschillen in begrijpend lezen. *Pedagogisch Tijdschrift*, 11 (2), 105-111.
- Aarnoutse, C.A.J. & Leeuwe, J.F.J. van (1988). Het belang van technisch lezen, woorden-schat en ruimtelijke intelligentie voor begrijpend lezen. *Pedagogische Studiën*, 65, 49-59.

- Aarnoutse, C.A.J., Leeuwe, J.F.J van, Oud, J.H.L. & Voeten, M.J.M. (1991). *Longitudinaal onderzoek schoolvorderingen* (subsidieaanvraag). Nijmegen: Vakgroep Onderwijskunde.
- Aarnoutse, C., Leeuwe, J. van, Voeten, R., Kan, N. van & Oud, H. (1996). *Longitudinaal onderzoek schoolvorderingen in het basisonderwijs*. Nijmegen: Vakgroep Onderwijskunde/Instituut voor Toepaste Sociale Wetenschappen.
- Aarnoutse, C., Leeuwe, J. van, Oude, H. Voeten, R. & Kan, N. van (1997). *Handleiding Nijmeegs leerlingvolgsysteem LISKAL*. Nijmegen: University Press.
- Aarnoutse, C.A.J., Mommers, M.J.C., Smits, B.W.G.M. & Leeuwe, J.F.J. van (1986). De ontwikkeling en samenhang van technisch lezen, begrijpend lezen en spellen. *Pedagogische Studiën*, 63, 97-110.
- Alexander, J.E. & Heathington, B.S. (1988). *Assessing and correcting classroom reading problems*. Glenview: Scott Foresman and Company.
- Allison, P.D. (1987). Estimation of linear models with incomplete data. In: C.C. Clogg (Ed.). *Sociological Methodology*, Vol. 17 (pp 71-103).
- Anderson, R.C. & Freebody, P. (1985). Vocabulary knowledge. Reprinted from Guthrie, 1981. In: H. Singer & R.B. Ruddell (Eds). *Theoretical models and processes of reading*, 3rd edition (pp 343-371). Newark, DE: International Reading Association.
- Arts, J., Filipiak, P. & Laarhoven, A. van (1992). *Analyse van methoden voor aanvankelijk lezen: Leeslijn*. 's-Hertogenbosch: KPC.
- Assenberg, M. (1992). *AVI-niveaulijsten voor het technisch lezen 1992*. 's-Hertogenbosch, 1992.
- Assenberg, M. & Schuurman-Hess, L.H. (1989). *AVI-niveaulijsten voor het technisch lezen*. 's-Hertogenbosch: KPC.
- Assink, E.M.H. (1989). Woordidentificatie bij leeszwakke kinderen. *Pedagogische Studiën*, 66 (10), 369-376.
- Assink, E.M.H. (1995). Use of visual, phonological and orthographic cues by poor readers. Paper presented at the 1995 Annual Meeting of the American Educational Research Association. San Francisco, April 18-22.
- Assink, E., Lam, M. & Knuijt, P. (1995). Gebruik van visuele, fonologische en orthografische informatie bij zwakke lezers. *Tijdschrift voor Taalbeheersing*, 17 (3), 192-206.
- Baar, K. de (samensteller en eindredacteur) (1992). *Leeslijn*. Amsterdam: Meulenhof Educatief.

- Barr, R. (1984). Beginning reading instruction. From debate to reformation. In: P.D. Pearson (Ed). *Handbook of reading research* (pp 545-581). New York: Longman.
- Beck, I.L. (1981). Reading problems and instructional practices. In: G.E. MacKinnon & T.G. Waller (Eds). *Reading research. Advances in theory and practice. Volume 2* (pp 53-95). New York: Academic Press.
- Beck, I.L. & McKeown, M.G. (1991). Conditions of vocabulary acquisition. In: R. Barr, M.L. Kamil, P.B. Mosenthal & P.D. Pearson (Eds). *Handbook of reading research. Volume II* (pp 789-814). Whiteplains, NY: Longman.
- Beck, I.L., Perfetti, C.A. & McKeown, M.G. (1982). The effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology*, 74, 506-521.
- Berg, R.M. van den (1972). *Werkmap individualisering van het leesonderwijs*. 's-Hertogenbosch: KPC.
- Berg, R.M. van den (1977). *Individualisering. Beschrijving en waardering van het onderwijsleergedrag bij individualiserend leesonderwijs in de lagere school*. Tilburg: Uitgeverij Zwijzen.
- Berg, R. van den (1991). Onderzoek naar het gebruik van het AVI-pakket. *Pedagogische Studiën*, 68 (5), 189-198.
- Berg, R.M. van den & Lintelo, H.G. te (1977a). *Observatie en begeleiding van individualiseringsvormen*. 's-Hertogenbosch: KPC.
- Berg, R.M. van den & Lintelo, H.G. te (1977b). *Individualisering van het leesonderwijs*. 's-Hertogenbosch: KPC.
- Boland, Th. (1991). *Lezen op termijn. Een onderzoek naar de ontwikkeling van de leesvaardigheid in het basisonderwijs en de invloed daarvan op de schoolloopbaan in het voortgezet onderwijs*. Proefschrift Nijmegen.
- Boland, Th. & Mommers, M.J.C. (1986). De ontwikkeling van het technisch lezen bij doorstromers en achterblijvers. In: P. Reitsma, A.G. Bus & W.H.J. van Bon (Red.). *Leren lezen en spellen: ontwikkeling en problemen* (pp 17-22). Lisse: Swets & Zeitlinger.
- Bollen, K.A. & Long, J.S. (Eds.) (1993). *Testing structural equation models*. Newbury Park, CA: Sage Publications Inc.

- Bon, W.H.J. van (1984). *Aspekten van taalvaardigheid. Een onderzoek naar de validiteit van de taaltests voor kinderen*. (Academisch proefschrift). Nijmegen: Katholieke Universiteit Nijmegen.
- Bon, W.H.J. van & Schreuder, R. (1986). Fonemische analyse: het effect van woordkenmerken. In: P. Reitsma, A.G. Bus & W.H.J. van Bon (Red.). *Leren lezen en spellen: ontwikkeling en problemen* (pp 24-31). Lisse: Swets & Zeitlinger.
- Bon, W.H.J. van, Tooren, P.H. & Eekelen, C.W.J.M. van (1995). Lexicale decisie en hardop lezen door normale en zwakke lezers. *Tijdschrift voor Taalbeheersing*, 17 (3), 181-191.
- Bos, K.P. van den (1989). Relationship between cognitive development, decoding skill, and reading comprehension in learning disabled Dutch children. In: P.G. Aaron & R.M. Joshi (Eds). *Reading and writing disorders in different orthographic systems*. Dordrecht: Kluwer Academic Publishers.
- Bos, K.P. van den, Lutje Spelberg, H.C., Scheepstra, A.J.M. & Vries, J.R. de (1994). *De Klepel*. Nijmegen: Berkhout Nijmegen BV.
- Bos, K.P. van den & Scheepstra, A.J.M. (1993). Het lezen van pseudowoorden en bestaande woorden. Deel II: decodeerroutes en decodeerproblemen. *Tijdschrift voor Orthopedagogiek*, 32, 225-237.
- Bos, K.P. van den, Scheepstra, A.J.M. & Lutje Spelberg, H.C. (1993). Het lezen van pseudowoorden en bestaande woorden. Deel I: De constructie en normering van de Klepel. *Tijdschrift voor Orthopedagogiek*, 32, 158-168.
- Broek, W. van den (1993). Theorieën van woordherkenning en praktische implicaties. *Tijdschrift voor Orthopedagogiek*, 32, 474-488.
- Brouwer, R.H.M. (1963). Onderzoek naar de leesmoeilijkheid van Nederlands proza. *Pedagogisch Studiën*, 40 (10), 454-464.
- Brouwer, R.H.M. (1976). Leesbaarheidsformules voor de Nederlandse taal. Een kritische beschouwing. *Pedagogisch Tijdschrift/Forum voor Opvoedkunde*, 1 (9), 491-505.
- Browne, M.W. & Cudeck, R. (1993). Alternative ways of assessing model fit. In: K.A. Bollen & J.S. Long (Eds.). *Testing structural equation models* (pp 136-162). Newbury Park, CA: Sage Publications Inc.
- Brus, B. Th. & Voeten, M.J.M. (1973). *Eén-Minuut-Test vorm A en B. Verantwoording en handleiding*. Nijmegen: Berkhout Testmateriaal BV.

- Bus, A.G. (1984). *Leesproblemen en instructiemethoden. Onderzoek naar voorbereidend, aanvankelijk en voortgezet leesonderwijs*. Lisse: Swets & Zeitlinger.
- Bus, A.G. (1990). Early reading acquisition. In: P. Reitsma & L. Verhoeven (Eds). *Acquisition of reading in Dutch* (pp 15-28). Dordrecht: Foris Publications.
- Bus, A.G. (1992). Ontluikende geletterdheid. In: L. Verhoeven (Red). *Handboek lees- en schrijfdidactiek. Functionele geletterdheid in basis- en voortgezet onderwijs* (pp 44-55). Amsterdam/Lisse: Swets & Zeitlinger.
- Bus, A.G., Both-de Vries, A., Buntinx, S. & Hardwarsing, S. (1994). Vier recent uitgebrachte leesmethoden voor 4- tot 8-jarigen in het licht van ontluikende geletterdheid. *Nederlands Tijdschrift voor Opvoeding, Vorming en Onderwijs*, 10 (1), 2-12.
- Calfee, R.C. & Hiebert, E. (1991). Classroom assessment of reading. In: R. Barr, M.L. Kamil, P.B. Mosenthal & P.D. Pearson (Eds). *Handbook of reading research. Volume II* (pp 281-309). Whiteplains, NY: Longman.
- Calfee, R.C. & Piontkowski, D.C. (1981). The reading diary: Acquisition of decoding. *Reading Research Quarterly*, 16 (3), 346-373.
- Caesar, F.B. (1982). *Vorbereidend en aanvankelijk leesonderwijs. Studie- en werkboek voor opleiding en bijscholing van leerkrachten*. Tilburg: Uitgeverij Zwijsen b.v.
- Chall, J.S. (1967). *Learning to read: The great debate*. New York: McGraw-Hill.
- Chall, J.S. (1983). *Stages of reading development*. New York: McGraw-Hill.
- Cipielewski, J. & Stanovich, K. E. (1992). Predicting growth in reading ability from children's exposure to print. *Journal of Experimental Child Psychology*, 54, 74-89.
- Cito (1979). *Lees en Begrijp 1. Handleiding en testmateriaal*. Arnhem: Cito.
- Cito (1980). *Lees en Begrijp 2. Handleiding en testmateriaal*. Arnhem: Cito.
- Cito (1981). *Begrijpend lezen bestemd voor leerjaar 3, 4 en 5 van het basisonderwijs. Handleiding en testmateriaal*. Arnhem: Cito.
- Clijnsen, A., Bruxvoort, R. van, Glauvé, M., Kooistra, R. & Leij, A. van der (1988). *Bronnenboek "Diagnostiek Lezen"*. Amsterdam: Vrije Universiteit.
- Corte, E. de, Geerligts, C.T., Lagerweij, N.A.J., Peters, J.J. & Vandenbergh, R. (1981). *Beknopte didaxologie*. Groningen: Wolters-Noordhoff.

- Cunningham, A.E. & Stanovich, K.E. (1990). Assessing print exposure and orthographic processing skill in children: A quick measure of reading experience. *Journal of Educational Psychology*, 82 (4), 733-740.
- Cunningham, A.E. & Stanovich, K.E. (1991). Tracking the unique effects of print exposure in children: Associations with vocabulary, general knowledge, and spelling. *Journal of Educational Psychology*, 83 (2), 264-274.
- Curtis, M.E. (1980). Development of components of reading skill. *Journal of Educational Psychology*, 72 (5), 656-669.
- Daneman, M. (1991). Individual differences in reading skills. In: R. Barr, M.L. Kamil, P.B. Mosenthal & P.D. Pearson (Eds). *Handbook of reading research. Volume II* (pp 512-538). Whiteplains, NY: Longman.
- Dickhout-Rutten, J.G.M. (1983). *De AVI-toetskaarten* (mededelingen nr. 13). Nijmegen: Vakgroep Interdisciplinaire Onderwijskunde.
- Dickhout-Rutten, A. & Mommers, M.J.C. (1984). AVI-niveaus gewikt en gewogen. *Jeugd in School en Wereld*, 69 (1), 11-13.
- Dommerholt, I. (1982). *Differentiële zinnenleestest*. Nijmegen: Berkhout Nijmegen BV.
- Dongen, D. van (1984). *Leesmoeilijkheden. Naar diagnostiserend onderwijzen bij het leren lezen*. Tilburg: Zwijsen b.v.
- Dongen, D. van, Hulsmans, H., Voeten, M.J.M. & Mommers, M.J.C. (1983). *Eindverslag van het longitudinale onderzoek (deelrapport 13, SVO-492)*. Nijmegen: Vakgroep Onderwijskunde.
- Doornbos, K. (1969). *Opstaan tegen het zittenblijven*. 's-Gravenhage: Staatsuitgeverij.
- Dymock, S. (1993). Reading but not understanding. *Journal of reading*, 37 (2), 86-91.
- Ehri, L.C. (1991). Development of the ability to read words. In: R. Barr, M.L. Kamil, P.B. Mosenthal & P.D. Pearson (Eds). *Handbook of reading research. Volume II* (pp 383-417). Whiteplains, NY: Longman.
- Ehri, L.C. & Wilce, L.S. (1979). Does word training increase or decrease interference in a Stroop task? *Journal of Experimental Child Psychology*, 27, 252-364.
- Ehri, L.C. & Wilce, L.S. (1983). Development of word identification speed in skilled and less skilled beginning readers. *Journal of Educational Psychology*, 75 (1), 3-18.

- Ehri, L.C. & Wilce, L.S. (1985). Movement into reading: Is the first stage of printed word learning visual or phonetic? *Reading Research Quarterly*, 20 (2), 163-179.
- Ehri, L.C. & Wilce, L.S. (1987). Cipher versus cue reading: An experiment in decoding acquisition. *Journal of Educational Psychology*, 79 (1), 3-13.
- Ellerman, H.H. & Spaaij, G.W.G. (1986). Inschaling van woorden naar moeilijkheidsgraad door beginnende lezers. In: P. Reitsma, A.G. Bus & W.H.J. van Bon (Red.). *Leren lezen en spellen: ontwikkeling en problemen* (pp 47-54). Lisse: Swets & Zeitlinger.
- Faber, S. (1989). De AVI-niveaus: Naar een genuanceerde benadering. *School & Begeleiding*, 6 (juni), 56-57.
- Faber, S. (1990). *Lezen op niveau*. Leiden: Schoolbegeleidingsdienst Rijnland.
- Faber, S. (1995). Het nieuwe AVI-toetspakket: een kritische bespreking. *School & Begeleiding*, 12 (3), 28-33.
- Farr, R. & Carey, R.F. (1986). *Reading. What can be measured?* Newark, DE: International Reading Association.
- Filipiak, P. & Laarhoven, A. van (1992a). *Analyse van methoden voor aanvankelijk lezen: De Leesbus*. 's-Hertogenbosch: KPC.
- Filipiak, P. & Laarhoven, A. van (1992b). *Analyse van methoden voor aanvankelijk lezen: De Leessleutel*. 's-Hertogenbosch: KPC.
- Filipiak, P. & Laarhoven, A. van & Winnubst, J. (1992). *Analyse van methoden voor aanvankelijk lezen: Veilig Leren Lezen*. 's-Hertogenbosch: KPC.
- Freeman-Smulders, A. (1995). Op school is lezen heel iets anders. *JSW*, 80 (1), 25-28.
- Geest, A.J.M. van der & Jorna, A. (1992). Leesonderwijs en leesprestaties. Consequenties voor leestheorie en leesdidactiek. *Tijdschrift voor Taalbeheersing*, 14 (2), 102-121.
- Gibson, E.J. (1972). Reading for some purpose. In: J.F. Kavanagh & I.G. Mattingly (Eds). *Language by ear and by eye* (pp 3-19). Cambridge, MA: MIT Press.
- Gibson, E.J. (1976). Learning to read. Reprinted from Science, 1965. In: H. Singer & R.B. Ruddell (Eds). *Theoretical models and processes of reading*, 2nd edition (pp 252-269). Newark, DE: International Reading Association.
- Gillijns, P. & Visser, J. (1995). De relatie tussen de Drie-Minuten-Toets en de AVI-toetskaarten. *School & Begeleiding*, 13 (2), 9-12.

- Goodman, K.S. (1973). *The psycholinguistic nature of the reading process, with a new foreword*. Detroit: Wayne State University Press.
- Goodman, K.S. (1976a). Behind the eye: What happens in reading. Reprinted from Goodman & Niles, 1970. In: H. Singer & R.B. Ruddell (Eds.). *Theoretical models and processes in reading, 2nd edition* (pp 470-496). Newark, DE: International Reading Association.
- Goodman, K.S. (1976b). Reading: A psycholinguistic guessing game. In: H. Singer & R.B. Ruddell (Eds.). *Theoretical models and processes in reading, 2nd edition* (pp 497-508). Newark, DE: International Reading Association.
- Goodman, K.S. (1985). A linguistic study of cues and miscues in reading. Reprint from Elementary English, 1965. In: H. Singer & R.B. Ruddell (Eds.). *Theoretical models and processes of reading, 3rd edition* (pp 129-134). Newark, DE: International Reading Association.
- Gough, P.B. (1972). One second of reading. In: J.F. Kavanagh & I.G. Mattingly (Eds.). *Language by ear and by eye* (pp 331-358). Cambridge, MA: MIT Press.
- Gough, P.B. (1983). Context, form, and interaction. In: K. Rayner (Ed). *Eye movements in reading: Perceptual and language processes* (pp 203-211). New York: Academic Press.
- Gough, P.B. (1984). Word recognition. In: P.D. Pearson (Ed). *Handbook of reading research* (pp 225-253). New York: Longman.
- Gough, P.B. & Hillinger, M.L. (1980). Learning to read: An unnatural act. *Bulletin of Orton Society*, 30, 179-196.
- Gough, P.B., Juel, C. & Roper-Schneider, D. (1983). A two-stage model of initial reading acquisition. In: J.A. Niles & L.A. Harris (Eds). *Searches for meaning in reading/language processing and instruction* (pp 207-211). Rochester, NY: National Reading Conference.
- Gough, P.B. & Tunmer, W.E. (1986). Decoding, reading and reading disability. *Remedial Special Education*, 7 (1), 6-10.
- Grainger, J. & Dijkstra, A.F.J. (1991). Het visuele woordherkenningsproces. In: A.J.W.M. Thomassen, L.G.M. Noordman & P.A.T.M. Eling (Red.). *Lezen en begrijpen. De psychologie van het leesproces* (p 47-68). Amsterdam/Lisse: Swets & Zeitlinger B.V.
- Groot, A.M.B. de (1991). Verschillen in leesvaardigheid. In: A.J.W.M. Thomassen, L.G.M. Noordman & P.A.T.M. Eling (Red.). *Lezen en begrijpen. De psychologie van het leesproces* (pp 201-222). Amsterdam/Lisse: Swets & Zeitlinger B.V.

- Harskamp, E.G., Deinum, J.F. & Lawant, C. (1995). Het gebruik en de resultaten van twee methoden voor aanvankelijk lezen. *School & Begeleiding*, 12 (4), 4-10.
- Hol, G.G.J.M., Haan, M. de & Kok, W.A.M. (1995). *De effectiviteit van methodes voor aanvankelijk leesonderwijs*. Utrecht: ISOR/Universiteit Utrecht.
- Hoover, W.A. & Gough, P.B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal*, 2 (2), 127-160.
- Horn, C.C. & Manis, F.R. (1987). Development of automatic and speeded reading of printed words. *Journal of Experimental Child Psychology*, 44 (1), 92-108.
- Hoyle, R.H. & Panter, A.T. (1995). Writing about structural equation models. In: R.H. Hoyle (Ed.). *Structural equation modelling: Concepts, issues, and applications* (pp 158-176). Thousand Oaks, CA: Sage Publications Inc.
- Hu, L.-T. & Bentler, P.M. (1995). Evaluating model fit. In: R.H. Hoyle (Ed.). *Structural equation modelling: Concepts, issues, and applications* (pp 76-99). Thousand Oaks, CA: Sage Publications Inc.
- Jansen, M.G.H. (1997). The Rasch Model for speed tests and some extensions with applications to incomplete designs. *Journal of Educational and Behavioral Statistics*, 22, summer issue.
- Johnson, D.D. & Baumann, J.F. (1984). Word identification. In: P.D. Pearson (Ed). *Handbook of reading research* (pp 583-608). New York: Longman.
- Johnston, P.H. (1984). Assessment in reading. In: P.D. Pearson (Ed). *Handbook of reading research* (pp 147-182). New York: Longman.
- Jöreskog, K.G. (1974). Analyzing psychological data by structural analysis of covariance matrices. In: D.H. Krantz, R.C. Atkinson, R.D. Luce & P. Suppes (Eds). *Contemporary developments in mathematical psychology. Vol II: Measurement, psychophysics, and neural information processing* (pp 1-56). San Francisco: W.H. Freeman and Company.
- Jöreskog, K.G. & Sörbom, D. (1993a). *Lisrel 8 user's reference guide*. Chicago: Scientific Software International Inc.
- Jöreskog, K.G. & Sörbom, D. (1993b). *Prelis 2 user's reference guide*. Chicago: Scientific Software International Inc.
- Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology*, 80 (4), 437-447.

- Juel, C. (1991). Beginning reading. In: R. Barr, M.L. Kamil, P.B. Mosenthal & P.D. Pearson (Eds). *Handbook of reading research. Volume II* (pp 759-788). Whiteplains, NY: Longman.
- Juel, C., Griffith, P.L. & Gough, P.B. (1985). Reading and spelling strategies of first-grade children. In: J.A. Niles & R.V. Lalik (Eds). *Issues in literacy: A research perspective* (pp 306-309). Rochester NY: National Reading Conference.
- Juel, C., Griffith, P.L. & Gough, P.B. (1986). Acquisition of literacy: A longitudinal study of children in first and second grade. *Journal of Educational Psychology*, 78 (4), 243-255.
- Just, M.A. & Carpenter, P.A. (1987). *The psychology of reading and language comprehension*. Newton: Allyn and Bacon.
- Klare, G.R. (1984). Readability. In: P.D. Pearson (Ed). *Handbook of reading research* (pp 681-744). New York: Longman.
- KPC, 1992a. *Instructie afname nieuwe AVI-toetsen voor Pabo-studenten*. 's-Hertogenbosch: KPC.
- KPC, 1992b. *Instructie afname herziene AVI-toetskaarten. Videoband*. 's-Hertogenbosch: KPC.
- Koornneef, M., Horst, A. van der & Aukes, N. (1994). *Registratie van leerlingkenmerken. Een inventarisatie van test- en toetsinstrumenten*. Hoevelaken: CPS.
- Krom, R. (1988). *Technisch lezen 3*. Arnhem: Cito.
- Kurvers, J. & Mooren, P. (1991). Het nieuwe leren lezen in de basisschool: een liaison dangereuse? *Leesgoed*, 4, 150-155.
- Laarhoven, A. van (1993). De nieuwe AVI-toets. *Trends*, 5 (4), 20-21.
- Laarhoven, A. van, Peterink, G. & Winnips, J. (1986). *Technisch lezen, een praktische uitwerking*. (2e gewijzigde druk). 's-Hertogenbosch, 1986.
- LaBerge, D. & Samuels, S.J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 6, 293-323.
- Lesgold, A.M. & Curtis, M.E. (1981). Learning to read words efficiently. In: A.M. Lesgold & C.A. Perfetti (Eds). *Interactive processes in reading* (pp 329-360). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

- Lesgold, A.M. & Resnick, L.B. (1982). How reading difficulties develop: Perspectives from a longitudinal study. In: J.P. Das, R. Mulcahey & A.E. Wall (Eds). *Theory and research in learning disabilities* (pp 155-187). New York: Plenum.
- Logan, G.D. (1988a). Automaticity, resources, and memory: theoretical controversies and practical implications. *Human Factors*, 30 (5), 583-598.
- Logan, G.D. (1988b). Toward an instance theory of automatization. *Psychological Review*, 95 (4), 492-527.
- Lord, F.M. & Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley Publishing Company.
- Lovett, M.W. (1987). A Developmental approach to reading disability: Accuracy and speed criteria of normal and deficient reading skill. *Child Development*, 58 (1), 234-260.
- Lundberg, I. (1991). Cognitive aspects of reading. *International Journal of Applied Linguistics*, 1 (2), 151-163.
- Malmquist, E. & Brus, B.Th. (1974). *Leren lezen, lezend leren*. Tilburg: Zwijsen.
- Manis, F.R. (1985). Acquisition of word identification skills in normal and disabled readers. *Journal of Educational Psychology*, 77, 78-90.
- Mason, J.M. (1980). When do children begin to read: An exploration of four year old children's letter and word reading competencies. *Reading Research Quarterly*, 15 (2), 203-227.
- McKeown, M.G., Beck, I.L., Omanson, R.C. & Pople, M.T. (1985). Some effects of the nature and frequency of vocabulary instruction on the knowledge and use of words. *Reading Research Quarterly*, 20, 522-535.
- Mezynski, K. (1983). Issues concerning the acquisition of knowledge: Effects of vocabulary training on reading comprehension. *Review of Educational Research*, 53, 253-279.
- Mommers, M.J.C. (1978). Hedendaagse theorieën over het leesproces en de structuur-methode voor het aanvankelijk lezen. *Pedagogische Studiën*, 55, 343-353;383-397.
- Mommers, M.J.C. (1983). *Caesar-één-minuut-test*. Tilburg: Uitgeverij Zwijsen BV.
- Mommers, M.J.C. (1987). An investigation into the relation between word recognition skills, reading comprehension and spelling skills in the first two years of primary school. *Journal of Research in Reading*, 10 (2), 122-143.

- Mommers, M.J.C. (1990). Metalinguistic awareness and learning to read. In: P. Reitsma & L. Verhoeven (Eds.). *Acquisition of reading in Dutch* (pp 29-42). Dordrecht: Foris Publications.
- Mommers, M.J.C., Leeuwe, J.F.J. van, Oud, J.H.L. & Janssens, J.M.A.M. (1986). Decoding skills, reading comprehension and spelling: A longitudinal investigation. *Tijdschrift voor Onderwijsresearch*, 11 (2), 97-113.
- Nagy, W.E., Herman, P.A. & Anderson, R.C. (1985). Learning words from context. *Reading Research Quarterly*, 20 (2), 233-253.
- Näslund, J.C. & Samuels, S.J. (1992). Automatic access to word sounds and meaning in decoding written text. *Reading and Writing Quarterly: Overcoming Learning Difficulties*, 8 (2), 135-156.
- Nathan, R.G. & Stanovich, K.E. (1991). The causes and consequences of differences in reading fluency. *Theory into Practice*, 30 (3), 176-184.
- Nicholson, T. (1986). Reading is not a guessing game: the great debate revisited. *Reading Psychology*, 7 (3), 197-210.
- Nicholson, T. (1992). Reading Wars: A brief history and an update. *International Journal of Disability, Development and Education*, 39 (3), 173-184.
- NICL (1989). *Toetsen en tests basisonderwijs*. Enschede: Nationaal InformatieCentrum Leermiddelen.
- NICL (1993). *Toetsen en tests aanvulling (incl. leerlingvolgsystemen) basisonderwijs*. Enschede: Nationaal InformatieCentrum Leermiddelen.
- Perfetti, C.A. (1984). Reading acquisition and beyond: Decoding includes cognition. *American Journal of Education*, 93 (1), 40-60.
- Perfetti, C.A. (1985). *Reading ability*. New York: Oxford University Press.
- Perfetti, C.A., Finger, E. & Hogaboam, T. (1978). Sources of vocalization latency differences between skilled and less skilled young readers. *Journal of Educational Psychology*, 70 (5), 730-739.
- Raftery, A.E. (1993). Bayesian model selection in structural equation modelling. In: K.A. Bollen & J.S. Long (Eds.). *Testing structural equation models* (pp 163-180). Newbury Park, CA: Sage Publications Inc.
- Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: University Press. (Oorspronkelijk gepubliceerd in 1960.)

- Rayner, K. & Pollatsek, A. (1989). *Psychology of reading*. Englewood Cliffs: Prentice Hall.
- Reicher, G.M. (1969). Perceptual recognition as a function of the meaningfulness of stimulus material. *Journal of Experimental Psychology*, 3, 513-518.
- Reitsma, P. (1983). *Phonemic and graphemic codes in learning to read*. (Academisch proefschrift). Amsterdam: Vrije Universiteit.
- Reitsma, P. (1988). Reading practice for beginners: Effects of guided reading, reading-while-listening and independent reading with computer-based speech feedback. *Reading Research Quarterly*, 23 (2), 219-235.
- Reitsma, P. (1990). Development of orthographic knowledge. In: P. Reitsma & L. Verhoeven (Eds). *Acquisition of reading in Dutch* (pp 43-64). Dordrecht: Foris Publications.
- Rovine, M.J. & Delaney, M. (1990). Missing data estimation in developmental research. In: A. von Eye (Ed.). *Statistical methods in longitudinal research. Vol I: Principles and structuring change* (pp 35-79). Boston: Academic Press.
- Samuels, S.J. (1979). The method of repeated readings. *The Reading Teacher*, 32, 403-408.
- Samuels, S.J. & Kamil, M.L. (1984). Models of the reading process. In: P.D. Pearson (Ed). *Handbook of reading research* (pp 185-224). New York: Longman.
- SBD Midden Holland & Rijnstreek (1992). *Bodegravense leestoets*.
- Scott, J.A. & Ehri, L.C. (1990). Sight word reading in prereaders: Use of logographic vs. alphabetic access routes. *Journal of Reading Behavior*, 22 (2), 149-166.
- Seegers, G.H.J. (1985). *Individuele verschillen in leesvaardigheid. Een onderzoek naar vaardigheidsverschillen in woordherkenning en de invloed hiervan op de prestaties op begrijpend lezen*. Nijmegen: Katholieke Universiteit.
- Seegers, G.H.J., Aarnoutse, C.A.J. & Mommers, M.J.C. (1987). *Lezen: een taalkundig-psychologische benadering*. Tilburg: Uitgeverij Zwijsen b.v.
- Simpson, G.B. & Lorsch, T.C. (1987). Automatic and conscious context effects in average and advanced readers. *Journal of Research in Reading*, 10 (2), 102-112.
- Smith, F. (1973). *Psycholinguistics and reading*. New York: Holt, Rinehart & Winston.
- Smith, F. (1979). *Reading without nonsense*. New York: Teachers College Press.

- Stanovich, K.E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, 16 (1), 32-71.
- Stanovich, K.E. (1982a). Individual differences in the cognitive processes of reading: I. Word decoding. *Journal of Learning Disabilities*, 15 (8), 485-493.
- Stanovich, K.E. (1982b). Individual differences in the cognitive processes of reading: II. Text-level processes. *Journal of Learning Disabilities*, 15 (9), 549-554.
- Stanovich, K.E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21 (4), 360-407.
- Stanovich, K.E. (1991). Word recognition: Changing perspectives. In: R. Barr, M.L. Kamil, P.B. Mosenthal & P.D. Pearson (Eds). *Handbook of reading research. Volume II* (pp 418-452). Whiteplains, NY: Longman.
- Stanovich, K.E. (1992). The psychology of reading: Evolutionary and revolutionary developments. *Annual Review of Applied Linguistics*, 12, 3-30.
- Stanovich, K.E., Cunningham, A.E. & Feeman, D.J. (1984). Relation between early reading acquisition and word decoding with and without context: A longitudinal study of first-grade children. *Journal of Educational Psychology*, 76 (4), 668-677.
- Stanovich, K.E. & West, R.F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly*, 24 (4), 402-433.
- Staphorsius, G. (1994). *Leesbaarheid en leesvaardigheid. De ontwikkeling van een domein-gericht meetinstrument*. Arnhem: Cito.
- Staphorsius, G. & Verhelst, N. (1997). Indexering van de leestechneek. *Pedagogische Studiën*.
- Stijnen, P.J.J. (1975). *Woordenschattest. Bestemd voor het derde tot en met zesde leerjaar van de basisschool. Verantwoording en handleiding*. Nijmegen: Berkhout BV.
- Struiksmā, A.J.C., Van der Leij, A. & Vieijra, J.P.M. (1995). *Diagnostiek van technisch lezen en aanvankelijk spellen. Zesde herziene druk* Amsterdam: VU Uitgeverij.
- Sulzby, E. & Teale, W. (1991). Emergent literacy. In: R. Barr, M.L. Kamil, P.B. Mosenthal & P.D. Pearson (Eds). *Handbook of reading research. Volume II* (pp 727-757). Whiteplains, NY: Longman.

- Szeszulski, P.A. & Manis, F.R. (1987). A comparison of word recognition processes in dyslexic and normal readers at two reading-age levels. *Journal of Experimental Child Psychology*, 44 (3), 364-376.
- Tanaka, J.S. (1993). Multifaceted conceptions of fit in structural equation models. In: K.A. Bollen & J.S. Long (Eds.). *Testing structural equation models* (pp 10-39). Newbury Park, CA: Sage Publications Inc.
- Theloosen, G. & Bon, W.H.J. van (1993). Herhaling van woorden en lettergroepen: oefeneffecten bij zwakke lezers. *Pedagogische Studiën*, 70, 180-194.
- Thomassen, A.J.W.M., Noordman, L.G.M. & Eling, P.A.T.M. (Red.) (1991). *Lezen en begrijpen. De psychologie van het leesproces*. Amsterdam/Lisse: Swets & Zeitlinger B.V.
- Touwen, R. (1990). *AVI-marktonderzoek. Een onderzoek naar het gebruik van het AVI-leespakket op basisscholen in Nederland*. 's-Hertogenbosch: KPC.
- Tunmer, W.E. & Chapman, J.W. (1995). Language prediction skill, phonological recoding ability, and beginning reading. Paper presented at the 1995 Annual Meeting of the American Educational Research Association. San Francisco, April 18-22.
- Tunmer, W.E. & Nesdale, A.R. (1985). Phonemic segmentation skill and beginning reading. *Journal of Educational Psychology*, 77 (4), 417-427.
- Venezky, R.L. (1984). The History of reading research. In: P.D. Pearson (Ed). *Handbook of reading research* (pp 3-38). New York: Longman.
- Verhallen, M. (1991) *Woordherkenning bij anderstalige kinderen. Stand van zaken in het basisonderwijs*. Amsterdam: ATW-publicatie nr. 57.
- Verhallen, M. (1994). *Lexicale vaardigheid van Turkse en Nederlandse kinderen: een vergelijkend onderzoek naar betekenisverlening*. Amsterdam: IFOTT.
- Verhallen, M. & Verhallen, S. (1991). *Woorden lezen, woorden vergelijken. Handreiking voor leraren in het basis- en voortgezet onderwijs*. Hoevelaken: CPS.
- Verhoeven, L. (1980a). *Technisch lezen 1*. Arnhem: Cito.
- Verhoeven, L. (1980b). *Technisch lezen 2*. Arnhem: Cito.
- Verhoeven, L.T. (1983). Diagnostische toetsen voor aanvankelijk lezen en spellen. *Pedagogische Studiën*, 60, 71-83.

- Verhoeven, L. (1992a). Aanvankelijk lezen en schrijven. In: Nijmeegse Werkgroep Taaldidactiek. *Taaldidactiek aan de basis. Vierde geheel herziene druk* (pp 241-274). Groningen: Wolters-Noordhoff.
- Verhoeven, L. (1992b). *Drie-Minuten-Toets*. Arnhem: Cito.
- Verhoeven, L. & Gillijns, P. (1994). Ontwikkelingen van beginnende lees- en spellingvaardigheid. *Tijdschrift voor Onderwijsresearch*, 19, 259-279.
- Verhoeven, L. & Mommers, C. (1989). De didactiek van het leren lezen: Van 'reading readiness' naar 'emergent literacy'. *School & Begeleiding*, 6 (21), 60-65.
- Verloop, N. & Schoot, F. van der (1995). Didactische evaluatie. In: J. Lowyck & N. Verloop. *Onderwijskunde. Een kennisbasis voor professionals* (pp 249-284). Groningen: Wolters-Noordhoff.
- Vos, T. de (1992). *Tempo test lezen (zinnen)*. Nijmegen: Berkhout Nijmegen BV.
- Vos, T. de (1993). *Tempo test lezen (woorden)*. Nijmegen: Berkhout Nijmegen BV.
- Visser, J., Laarhoven, A. van & Beek, A. ter (1994). *AVI-toetspakket. Handleiding*. 's-Hertogenbosch: KPC.
- Visser, J., Laarhoven, A. van & Beek, A. ter (1996). *AVI-toetspakket. Handleiding. Derde herziene druk*. 's-Hertogenbosch: KPC.
- Visser, J. & Voeten, M. (1994). *Schaling en normering van de nieuwe AVI-toetskaarten*. Paper voor de ORD 1994 te Utrecht. 's-Hertogenbosch: KPC.
- Wiegersma, S. (1986). *Leesvaardigheidstest voor het onderzoek van de mechanische leesvaardigheid*. Lisse: Swets & Zeitlinger BV.
- Wassenaar, W. (1989). *Diagnostische toetsbatterij technische leesvaardigheid*. Nijmegen: Berkhout Nijmegen BV.
- Yap, R. & Leij, A. van der (1995). Decodeerproblemen. *Tijdschrift voor Taalbeheersing*, 17 (3), 164-180.
- Zondervan, F., Steen, P. van & Gunneweg, G. (1976). De leesbaarheid van basisschoolteksten. Objectieve ordeningscriteria voor instructieve teksten. *De Nieuwe Taalgids*, 69, 426-445.

BIJLAGE 1 OVERZICHT VAN DE AVI-NIVEAUS

De negen AVI-niveaus, en dus ook de bijbehorende toetskaarten, onderscheiden zich van elkaar op woord- en zins- en tekstkenmerken en de Leesindex A van Brouwer (1963). Deze index is gebaseerd op een formule (B1.1) waarin de woordlengte (gemiddelde aantal lettergrepen per woord) en de zinslengte (gemiddelde aantal woorden per zin) van de tekst zijn opgenomen. Hoe hoger het gemiddelde aantal lettergrepen per woord is en/of hoe hoger het gemiddelde aantal woorden per zin, des te moeilijker de tekst is. De formule is zodanig geconstrueerd dat een lagere waarde van de index wijst op een moeilijker tekst.

$$\text{Leesindex A} = 195 - (2/3) * \text{WL} - 2 * \text{ZL} \quad (\text{B1.1})$$

waarbij :

WL = $100 * \text{aantal lettergrepen} / \text{aantal woorden}$

ZL = $\text{aantal woorden} / \text{aantal zinnen}$

AVI-niveau 1

Tekstkenmerken:

- korte zinnen, één zin per regel;
- samengestelde zinnen over twee regels verdeeld, komen voor (voor leesindex geteld als twee zinnen);
- hoofdletters kunnen voorkomen;
- leesindex A: > 100.

Woordtypen:

- éénlettergrepige woorden:
 - . medeklinker-klinker: zo, pa;
 - . klinker-medeklinker: ik, om;
 - . medeklinker-klinker-medeklinker: naam, roos;
- éénlettergrepige woorden met één medeklinkercombinatie kunnen voorkomen.

AVI-niveau 2

Tekstkenmerken:

- hoofdletters kunnen voorkomen;
- korte zinnen, één zin per regel;
- samengestelde zinnen worden nog per zin over twee regels verdeeld (voor de leesindex geteld als twee zinnen);
- leesindex A: > 100.

Woordtypen (uitbreiding):

- éénlettergrepige woorden:
 - . eindigend op dt, ng of nk: houdt, bang, zink;
 - . twee medeklinkers voor en/of achteraan het woord: stoel, pats, stelt;
 - . drie medeklinkers voor of achteraan het woord: straal, kempt;
- tweelettergrepige of samengestelde woorden zonder spellingsmoeilijkheden: voetbal, mamma;
- verkleinwoorden: huisje.

AVI-niveau 3

Tekstkenmerken:

- korte zinnen kunnen betekenisvol zijn afgebroken en doorlopen op de volgende regel. De nieuwe zin begint nog wel op een nieuwe regel;
- leesindex A: > 100.

Woordtypen (uitbreiding):

- alle typen éénlettergrepige woorden;
- alle typen tweelettergrepige woorden;
- drie- en meerlettergrepige woorden zonder spellingsmoeilijkheden.

AVI-niveau 4

Tekstkenmerken:

- zinnen kunnen, betekenisvol afgebroken, doorlopen op de volgende regel;
- een gesproken zin begint nog steeds op een nieuwe regel;
- leesindex A: > 100.

Woordtypen (uitbreiding):

- alle typen drie- en meerlettergrepige woorden;
- leenwoorden zonder lastige, afwijkende teken-klankcombinaties: portemonnee, disette.

AVI-niveau 5 (vanaf dit niveau zijn alleen tekstkenmerken bepalend)

- leesindex A: 99-94;
- gemiddelde zinslengte: 7 woorden (6-8 woorden);
- gemiddelde woordlengte: 127 lettergrepen (123-130).

AVI-niveau 6

- leesindex: A 93-89;
- gemiddelde zinslengte: 8 woorden (7-9 woorden);
- gemiddelde woordlengte: 132 lettergrepen (129-135).

AVI-niveau 7

- leesindex A: 88-84;
- gemiddelde zinslengte: 9 woorden (8-10);
- gemiddelde woordlengte: 136 lettergrepen (134-139).

AVI-niveau 8

- leesindex A: 83-79;
- gemiddelde zinslengte: 10 woorden (9-11);
- gemiddelde woordlengte: 141 lettergrepen (138-144):

AVI-niveau 9

- leesindex A: 78-74;
- gemiddelde zinslengte: 11 woorden (10-12);
- gemiddelde woordlengte: 146 lettergrepen (143-148):

niveau 1A

naar het bos

pa leest een boek.
ma zit in de tuin.
en wat doet tom?
hij kijkt uit het raam.

wat is het hier saai, roept tom.
ik wil naar het bos.
pap, ga je mee?
nee, ik lees al, zegt pa.

tom loopt naar de tuin.
mam, ga je mee naar het bos?
ja tom, dat is leuk.

ma komt het huis in.
ze geeft pa een duw.
ga jij ook mee? zegt ze.
pa kijkt wat sip.

dan rent pa naar de hal.
hij pakt zijn pet.
ik ga naar het bos, roept hij.
het is hier veel te saai.
wie gaat er mee?

De lucht in

Stel je eens voor dat je een reis mag maken. Het wordt geen gewone reis. Je gaat niet met de trein, de auto of het vliegtuig. Nee, je gaat met een luchtballon.

Als je met je verjaardag zo'n reis mag maken, zou je dan meegaan? Zou je het aandurven om je woonplaats van grote hoogte te bekijken?

Mensen met hoogtevrees kunnen beter thuis blijven. Maar jij durft vast wel met een echte ballonvaarder de tocht te maken.

Natuurlijk wil je wel eerst weten hoe de ballon opstijgt en in de lucht blijft. En ook hoe sterk het materiaal is waarvan de ballon gemaakt wordt. Als je daarover informatie wilt hebben, dan kun je in iedere bibliotheek voldoende boeken vinden.

De werking komt in het kort hierop neer. Precies onder de opening van de luchtballon zijn grote gasbranders bevestigd. Die verwarmen de lucht. En dat is belangrijk, want als lucht verwarmd wordt, stijgt ze op. Het duurt wel even voordat de grote ballon gevuld is met hete lucht.

Bij het opstijgen maken de branders nog steeds veel lawaai, omdat de lucht verder op temperatuur moet komen. Maar wanneer je hoog in de lucht zweeft, branden ze alleen als de ballon te veel daalt. Als beloning voor je moed kun je uiteindelijk in alle stilte genieten van het uitzicht. Wat ziet de wereld er van boven mooi uit!

BIJLAGE 3 OORSPRONKELIJKE INDELING AVI-NIVEAUS

De oorspronkelijke indeling van de AVI-niveaus (Van den Berg & Te Lintelo, 1977b):

AVI-niveau 1

- éénlettergrepige woorden waarbij zelden medeklinkerverbindingen aan begin of eind van een woord voorkomen (sch aan begin woord geldt als één medeklinker). Dus woorden van het type:
 - . klinker-medeklinker (en, op, aap, oom)
 - . medeklinker-klinker (na, zo);
 - . medeklinker-klinker-medeklinker (boom, mus vuur, bel);
 - . twee medeklinkers aan eind van woord mag, mits de eerste van die twee geen 'd' is (nest, loopt);
- het tweelettergrepige woord 'moeder';
- korte zinnen, één zin per regel;
- geen hoofdletters;
- leesindex A: 110 en hoger.

AVI-niveau 2

- zie niveau 1;
- éénlettergrepige woorden
 - . met 'd' en 'b' aan het eind (koud, web);
 - . persoons-t achter stam na twee medeklinkers, uitgezonderd als de tweede medeklinker een 'd' is (harkt, pompt);
 - . met 'ng' en 'nk' aan eind (wang, bang, bank);
 - . met dubbele medeklinker aan begin woord (klein, stoel, ster);
- tweelettergrepige woorden zonder spellingsmoeilijkheden (hakken, parken);
- het woord 'vader';
- verkleinwoorden (poesje, doosje)
- korte zinnen, één zin per regel
- geen hoofdletters;
- leesindex A: 110 en hoger.

AVI-niveau 3

- zie niveau 2;
- tweelettergrepige woorden met klinker in open lettergreep (tafel, komen, eten);
- persoons-t achter stam met d (houdt, wordt);
- woorden met drie medeklinkers aan het begin (springt) of vier aan het eind (hengst, vangst, angst);
- drie- en meerlettergrepige woorden zonder spellingsmoeilijkheden (boerderij, kinderen);
- woorden met voorvoegsels ge-, be- en ver- (verhaal, gebracht);
- korte zinnen die kunnen overlopen;
- hoofdletters;
- leesindex A 109-100.

AVI-niveau 4

- zie niveau 3;
- woorden eindigend op -lijk, -ig, enz. (gelukkig, heerlijk, gevaarlijk);
- andere drie- en meerlettergrepige woorden (muzikanten, marionetten);
- zowel korte hoofdzin + korte bijzin als andersom;
- overlopende zinnen (wel korte zinnen);
- leesindex A 109-100.

AVI-niveau 5

- zie niveau 4;
- leesindex A 99-94;
- gemiddelde woordlengte 127 lettergrepen (123-130);
- gemiddelde zinslengte 7 woorden (6-8 woorden).

AVI-niveau 6

- zie niveau 5;
- leesindex A 93-89;
- gemiddelde woordlengte 132 lettergrepen (129-135);
- gemiddelde zinslengte 8 woorden (7-9 woorden).

AVI-niveau 7

- zie niveau 6;
- leesindex A 88-84;
- gemiddelde woordlengte 136 lettergrepen (134-139);
- gemiddelde zinslengte 9 woorden (8-10).

AVI-niveau 8

- zie niveau 7;
- leesindex A 83-79;
- gemiddelde woordlengte 141 lettergrepen (138-144);
- gemiddelde zinslengte 10 woorden (9-11).

AVI-niveau 9

- zie niveau 8;
- leesindex A 78-74;
- gemiddelde woordlengte 146 lettergrepen (143-148);
- gemiddelde zinslengte 11 woorden (10-12).

BIJLAGE 4 KORTE INSTRUCTIE TRY-OUT JANUARI 1992

- Leg alle spullen klaar:
 - leerlinglijsten voor de negen AVI-niveaus
 - de oude AVI-kaarten, versie A
 - de normering van de oude AVI-kaarten, versie A
 - de experimentele AVI-toetskaarten, versie A en B
 - scoringsformulieren
 - potlood
 - stopwatch
- Vul de logistieke gegevens in op de scoringsformulieren:
 - de datum
 - het schoolnummer
 - het leerlingnummer
 - het groepsnummer
 - het meetmoment
- Bepaal het AVI-niveau van de leerling door de oude AVI-toetskaarten (versie A) af te nemen
- Noteer op de leerlinglijst voor het specifieke niveau:
 - de naam en het nummer van de leerling
 - de leestijd en het aantal fouten op de oude kaart
- Start met de eerste experimentele AVI-toetskaart op de leerlinglijst
- Bespreek kort de tekening
- Zeg:

Ik heb hier een verhaaltje voor je. Probeer jij nu dit verhaaltje zo goed mogelijk voor te lezen. Je moet proberen het verhaal zo vlot mogelijk en met zo weinig mogelijk fouten te lezen. Ik lees de titel voor en dan mag jij beginnen.

Wanneer je merkt dat je een woord fout hebt gelezen, hoef je alleen dat woord te verbeteren en mag je gewoon weer verder lezen. Je hoeft niet aan het begin van de zin te beginnen.

Ik lees eerst de titel voor en dan lees jij het verhaaltje verder.
- Start de stopwatch
- Noteer de fouten op het scoringsformulier
- Als een kind langer dan 5 seconden over een woord doet: voorzeggan.
- Wanneer het kind klaar is, drukt u de stopwatch in. Noteer de leestijd en het aantal zwart gemaakte hokjes op het scoringsformulier
- Neem op dezelfde wijze de volgende twee kaarten op de lijst af
- Stop de afname na drie kaarten
- Neem de tweede serie kaarten minimaal na één dag, maar binnen een week na de eerste afname af.

Wat is fout:

- een verkeerd woord lezen, zonder het alsnog te verbeteren (F)
- een woord spellend lezen, zonder het daarna alsnog in een keer te lezen (S)
- een woord overslaan (O)
- woorden omwisselen (O)
- woorden tussenvoegen (T)
- als het oplezen van een woord langer dan 5 seconden duurt. In dat geval wordt het woord voorgezegd(V)

Wat is niet fout:

- verkeerde klemtoon
- versprekingen en haperingen, waarna het woord alsnog goed wordt gelezen
- een woord wordt goed gelezen, nadat het eerst gespeld is
- een fout woord wordt verbeterd door het goede woord
- dialectuitspraak

'De hond van Ans is niet bang', Niveau 3 A pag. 1

Datum

Schoolnummer

10 20 30 40 50 60 70 80 90
0 1 2 3 4 5 6 7 8 9

Groepnummer

3 4 5 6 7 8

Aantal streepjes:

10 20 30 40 50 60 70 80 90
0 1 2 3 4 5 6 7 8 9

Leerlingnummer

100 200 300 400 500 600 700 800 900
10 20 30 40 50 60 70 80 90
0 1 2 3 4 5 6 7 8 9

Meetmoment

1 2 3 4 5 6 7 8 9

Kaartnr A of B en pagina 1 of 2

1 2 3 4 5 6 7 8 9
A B 1 2

Leestijd in minuten

0 1 2 3 4 5 6 7 8 9

Leestijd in seconden

10 20 30 40 50 60 70 80 90
0 1 2 3 4 5 6 7 8 9

	F	S	O	T	V		F	S	O	T	V		F	S	O	T	V
Ans						bomen						Tim					
heeft						of						spraak					
een						doet						naar					
lieve						een						de					
hond						plas						Duitse					
Hij						Na						herder					
heeft						een						Die					
zwart						poosje						schrok					
haar						trekt						heel					
en						Ans						erg					
heet						even						en					
Tim						aan						begon					
Het						de						te					
is						riem						janken					
een						Laatst						Doe					
poedel						kwam						die					
Elke						ze						hond					
dag						Liesbet						weg,					
laat						tegen						nep					
Ans						Zij						Liesbet					
hem						heeft						Hier					
uit						ook						Tim,					
Dan						een						kom					
gaat						hond						bij					
ze						Dat						de					
naar						is						beas,					
het						een						nep					
park						grote						Ans					
Samen						Duitse						Maar					
lopen						herder						het					
ze						Tim						geblaf					
het						blafte						hield					
park						naar						niet					
een						hem						op					
keer						en						Ans					
rond						trok						pakte					
Soms						aan						Tim					
staat						de						bij					
Tim						riem						de					
stil						Ans						riem					
Hij						liet						en					
ruikt						de						ging					
dan						riem						naar					
aan						los						huis					

BIJLAGE 6 DESCRIPTIEVE STATISTIEKEN PER TOETSKAART IN DE TRYOUT JANUARI 1992

Tabel B6.1: Statistische gegevens leestijden per toetskaart

	Gemiddelde	Standaard- afwijking	min	max	N	aantal woorden per minuut
kaart1A	153,1	82,9	38	433	139	40,8
kaart1B	167,0	85,6	44	455	137	37,4
kaart2A	161,8	91,3	56	540	160	47,8
kaart2B	158,5	94,4	49	556	154	46,9
kaart3A	128,9	51,1	58	295	141	61,4
kaart3B	135,2	49,6	60	300	137	63,5
kaart4A	118,0	35,1	68	249	127	75,8
kaart4B	110,1	31,3	64	241	124	85,0
kaart5A	105,8	25,1	50	210	116	93,0
kaart5B	106,2	24,5	65	198	113	87,0
kaart6A	83,6	24,0	55	307	129	107,7
kaart6B	85,3	16,3	52	168	129	106,2
kaart7A	113,9	18,9	69	173	181	108,5
kaart7B	108,9	20,7	64	181	180	109,6
kaart8A	98,4	21,5	56	180	291	124,4
kaart8B	82,5	17,6	46	145	289	132,4
kaart9A	106,0	17,6	65	160	225	130,2
kaart9B	105,8	19,0	68	206	224	128,2

Tabel B6.2: Statistische gegevens aantal fouten per toetskaart

	Gemiddelde	Standaard- afwijking	min	max	N	percentage fouten
kaart1A	8,9	11,0	0	56	139	8,6
kaart1B	9,1	10,8	0	51	137	8,8
kaart2A	6,3	7,8	0	42	160	4,9
kaart2B	5,5	8,1	0	51	154	4,4
kaart3A	4,7	4,5	0	20	141	3,6
kaart3B	5,2	4,3	0	20	137	3,6
kaart4A	4,5	4,0	0	22	127	3,0
kaart4B	3,4	3,4	0	17	124	2,2
kaart5A	2,1	2,1	0	11	116	1,3
kaart5B	3,8	3,2	0	19	113	2,5
kaart6A	2,0	2,1	0	9	129	1,3
kaart6B	2,1	2,4	0	19	129	1,4
kaart7A	3,7	3,1	0	19	181	1,8
kaart7B	3,8	3,5	0	18	180	1,9
kaart8A	3,7	3,6	0	23	291	1,8
kaart8B	2,9	2,6	0	15	289	1,6
kaart9A	3,2	3,3	0	20	225	1,4
kaart9B	3,3	2,8	0	14	224	1,5

BIJLAGE 7 KORTE INSTRUCTIE NORMERINGSONDERZOEK

- Leg alle spullen klaar:
 - de AVI-toetskaarten;
 - scoreformulieren;
 - potlood + gum;
 - eenvoudige stopwatch.
- Vul de logistieke gegevens in op de scoreformulieren:
 - de datum;
 - de naam en de geboortedatum van de leerling;
 - het groepsnummer;
 - naam van degene die de toets afneemt.
- Start met de afname van de eerste AVI-toetskaart.
- Bespreek kort de tekening.
- Zeg:

Ik heb hier een verhaaltje voor je. Probeer jij nu dit verhaaltje zo goed mogelijk voor te lezen. Je moet proberen het verhaal zo vlot mogelijk en met zo weinig mogelijk fouten te lezen. Ik lees de titel voor en dan mag jij beginnen.

Wanneer je merkt dat je een woord fout hebt gelezen, hoef je alleen dat woord te verbeteren en mag je gewoon weer verder lezen. Je hoeft niet aan het begin van de zin te beginnen.

Ik lees eerst de titel voor en dan lees jij het verhaaltje verder.
- Start de stopwatch.
- Noteer de fouten op het scoreformulier.
- Als een kind langer dan 5 seconden over een woord doet: voorzeggen.
- Wanneer het kind klaar is, drukt u de stopwatch in. Noteer de leestijd en het aantal fouten op het scoreformulier.
- Zoek in de normtabel op wat de prestatie van de leerling op de kaart is:
 - bij goed: neem de kaart van één niveau hoger af;
 - bij voldoende: het kind leest op het AVI-niveau dat correspondeert met het nummer van de afgenomen toetskaart;
 - bij onvoldoende: neem de kaart van één niveau lager af.

Wat is fout:

- een verkeerd woord lezen, zonder het alsnog te verbeteren (F)
- een woord spellend lezen, zonder het daarna alsnog in een keer te lezen (S)
- een woord overslaan (O)
- woorden omwisselen (O)
- woorden tussenvoegen (T)
- als het oplezen van een woord langer dan 5 seconden duurt. In dat geval wordt het woord voorgezegd(V)

Wat is niet fout:

- verkeerde klemtoon
- versprekingen en haperingen, waarna het woord alsnog goed wordt gelezen
- een woord wordt goed gelezen, nadat het eerst gespeld is
- een fout woord wordt verbeterd door het goede woord
- dialectuitspraak
- ze i.p.v. zij, je i.p.v. jij, we i.p.v. wij, z'n i.p.v. zijn

BIJLAGE 8 OMZETTINGSTABELLEN TEN BEHOEVE VAN TOEWIJZEN TOETS- KAARTEN

Onderstaande tabellen zijn gebruikt om een inschatting te maken van het leesniveau van de leerlingen ten behoeve van het toewijzen van de toetskaarten in het onderzoek. Onderstaande tabel (B8.1) gebruikt u om de leerlingen van groep 4 en 5 in te delen naar AVI-niveau.

Tabel B8.1: Omzettingstabel Eén-Minuuut-Test naar AVI-niveau

<i>EMT (Brus)</i> (aantal goed)	AVI-niveau
0 - 22	1
23 - 29	2
30 - 35	3
36 - 40	4
41 - 46	5
47 - 51	6
52 - 57	7
58 - 65	8
66 of hoger	9

Nota Bene:

Leerlingen die een EMT-score hebben van 0 tot en met 16, zullen feitelijk nog niet op AVI-niveau 1 lezen (soms AVI-0 genoemd). In het onderzoek krijgen deze leerlingen echter dezelfde kaarten te lezen als de leerlingen die wel op AVI-niveau 1 lezen. Deze leerlingen worden door u dan ook op de leerlinglijst van 'AVI-niveau 1' bijgeschreven, en krijgen de kaarten te lezen die op die lijst staan aangegeven.

Onderstaande tabel (B8.2) kan gebruikt worden voor groep 3, wanneer u de Caesar-EMT afneemt.

Tabel B8.2: Omzettingstabel Caesar-EMT naar AVI-niveau

<i>Caesar-EMT</i> (aantal goed)	AVI-niveau
0 - 40	1
41 - 51	2
52 - 67	3
68 of meer	4

Ook nu geldt dat een groep leerlingen feitelijk nog niet AVI-niveau 1 zal lezen, namelijk de leerling met een score op de Caesar-EMT van 0 tot en met 31.

BIJLAGE 9 DESCRIPTIEVE STATISTIEKEN PER TOETSKAART IN HET NORMERINGSONDERZOEK

De volgende tabellen bevatten de beschrijvende statistieken van de leestijden en fouten per toetskaart. De tabel B9.1 en B9.2 bevatten per groep de leestijden respectievelijk het aantal fouten gemeten in maart 1992. De tabellen B9.3 tot en met B9.6 bevatten deze gegevens voor oktober 1992 respectievelijk maart 1993. Het aantal leerlingen dat een toetskaart heeft gelezen kan verschillen wanneer we de tabel voor de leestijden vergelijken met die voor het aantal fouten als gevolg van ontbrekende scores.

Tabel B9.1: Leestijd per groep op AVI-toetskaarten gemeten in maart 1992 (gemiddelde, standaardafwijking, minimum en maximum)

toetskaart	groep 3					groep 4					groep 5				
	Gemid- delde	Standaard- afwijking	min	max	N	Gemid- delde	Standaard- afwijking	min	max	N	Gemid- delde	Standaard- afwijking	min	max	N
1A	140,8	75,2	43	630	951	80,6	21,6	49	156	198	68,4	10,3	59	90	8
1B	154,3	79,4	33	592	929	92,1	28,5	45	271	202	82,3	9,7	65	96	8
2A	180,9	94,2	40	596	880	98,0	31,5	53	259	321	86,6	16,3	57	148	29
2B	177,8	91,8	45	609	882	95,2	33,2	52	285	321	79,7	19,0	49	142	27
3A	154,0	79,4	48	552	212	106,0	27,2	46	236	383	94,9	17,5	64	152	83
3B	156,1	74,2	51	478	209	112,4	29,2	57	287	382	101,5	21,5	64	182	83
4A	124,8	52,5	57	322	80	105,7	24,0	58	290	407	100,5	22,8	53	214	154
4B	122,9	59,7	52	344	76	98,3	21,6	59	290	400	94,0	19,8	53	175	155
5A	130,2	75,5	65	378	30	102,8	20,9	47	182	376	98,7	18,1	60	201	245
5B	128,5	70,7	60	346	30	99,6	20,4	60	172	377	99,0	20,2	54	190	250
6A	92,2	25,1	62	153	12	83,6	15,3	55	141	378	81,0	14,0	45	158	368
6B	92,3	19,3	61	116	11	84,1	14,9	42	159	377	82,9	15,4	50	202	371
7A	146,6	43,7	101	196	5	117,6	19,9	77	240	352	114,4	19,7	69	193	492
7B	134,4	27,4	101	159	5	110,7	20,5	68	240	346	109,6	18,8	70	187	496
8A	112,0	7,1	107	117	2	108,8	18,6	65	177	324	102,5	19,3	53	166	651
8B	94,5	9,2	88	101	2	91,8	15,0	55	150	322	87,4	16,8	50	167	649
9A	134,5	14,8	124	145	2	118,5	18,8	72	201	190	111,1	19,5	61	185	472
9B	135,5	4,9	132	139	2	118,5	18,8	59	191	189	111,2	19,3	63	233	468

Tabel B9.2: Aantal fouten per groep op AVI-toetskaarten gemeten in maart 1992 (gemiddelde, standaardafwijking, minimum en maximum)

toetskaart	groep 3					groep 4					groep 5				
	Gemid- delde	Standaard- afwijking	min	max	N	Gemid- delde	Standaard- afwijking	min	max	N	Gemid- delde	Standaard- afwijking	min	max	N
1A	5,3	7,1	0	80	953	2,8	2,6	0	13	198	2,5	2,3	0	7	8
1B	5,2	7,3	0	97	932	3,0	2,6	0	13	202	2,6	4,8	0	14	8
2A	5,9	7,8	0	57	890	2,9	2,8	0	17	322	2,7	3,0	0	17	29
2B	5,4	7,4	0	55	890	2,4	2,5	0	19	321	1,5	1,8	0	8	27
3A	5,4	5,8	0	35	212	4,2	3,6	0	20	384	3,0	3,0	0	17	83
3B	5,7	5,3	0	42	209	4,9	3,9	0	24	382	3,6	3,4	0	19	84
4A	3,1	3,2	0	19	80	4,4	3,6	0	20	407	3,6	3,0	0	14	155
4B	3,4	4,1	0	16	76	3,2	2,9	0	17	400	2,7	2,3	0	11	157
5A	2,5	3,1	0	14	31	2,8	2,6	0	19	379	2,2	1,9	0	11	245
5B	2,4	3,0	0	13	30	3,1	2,9	0	16	377	2,8	2,5	0	15	250
6A	2,3	2,9	0	11	12	2,0	2,0	0	9	380	1,8	1,8	0	11	368
6B	1,1	1,4	0	4	11	2,1	2,3	0	12	378	2,0	1,9	0	13	372
7A	6,4	6,1	2	17	5	4,0	3,7	0	36	352	3,7	3,0	0	20	492
7B	4,4	6,0	0	15	5	4,2	3,5	0	20	347	3,9	3,3	0	25	496
8A	6,0	1,4	5	7	2	4,0	3,3	0	18	327	3,6	3,1	0	25	652
8B	2,0	1,4	1	3	2	3,6	3,0	0	16	322	3,2	2,6	0	18	649
9A	7,0	1,4	6	8	2	4,3	3,7	0	25	190	4,1	3,5	0	25	472
9B	6,0	2,8	4	8	2	4,2	3,5	0	19	189	3,8	3,3	0	18	468

Tabel B9.3: Leestijd per groep op AVI-toetskaarten gemeten in oktober 1992 (gemiddelde, standaardafwijking, minimum en maximum)

toetskaart	groep 4					groep 5					groep 6				
	Gemid- delde	Standaard- afwijking	min	max	N	Gemid- delde	Standaard- afwijking	min	max	N	Gemid- delde	Standaard- afwijking	min	max	N
1A	86,7	36,3	36	249	521	79,6	21,5	48	128	25	79,0	10,2	66	89	4
2A	99,7	44,4	25	317	796	88,2	28,2	50	230	120	92,1	17,0	67	131	20
3A	107,3	44,2	20	300	622	98,4	33,2	54	284	216	97,6	19,0	60	159	72
4A	103,9	36,4	48	358	404	102,0	29,9	49	217	328	97,8	19,9	64	176	147
5A	91,7	19,9	57	147	144	97,3	21,1	50	200	358	81,9	17,3	55	187	254
6A	78,3	15,9	47	139	88	81,0	17,0	46	170	374	111,7	19,3	70	199	376
7A	109,8	18,6	75	156	48	113,9	20,5	73	223	363	97,1	19,7	55	184	644
8A	110,3	24,8	75	168	15	100,9	17,5	61	176	367	105,3	19,5	55	196	533
9A	103,0	16,2	80	118	5	109,4	17,2	65	172	258					

Tabel B9.4: Aantal fouten per groep op AVI-toetskaarten gemeten in oktober 1992 (gemiddelde, standaardafwijking, minimum en maximum)

toetskaart	groep 4				N	groep 5				N	groep 6				N
	Gemid- delde	Standaard- afwijking	min	max		Gemid- delde	Standaard- afwijking	min	max		Gemid- delde	Standaard- afwijking	min	max	
1A	2,9	3,3	0	24	521	3,3	3,3	0	11	25	3,0	3,6	0	8	4
2A	2,6	3,3	0	36	797	2,9	2,4	0	9	120	5,7	4,5	0	20	20
3A	4,1	3,8	0	25	622	4,4	3,9	0	29	216	4,9	4,3	0	24	72
4A	3,8	3,4	0	23	406	4,5	3,7	0	28	328	3,4	3,3	0	27	147
5A	2,0	2,1	0	11	144	2,6	2,2	0	11	359	2,4	2,3	0	14	254
6A	1,5	1,6	0	7	88	2,4	2,5	0	20	374	4,0	3,0	0	16	377
7A	2,6	2,1	0	9	48	4,3	3,5	0	19	364	3,5	3,2	0	23	644
8A	3,8	2,5	0	9	15	3,9	3,2	0	19	369	3,6	3,6	0	49	533
9A	3,4	2,1	1	6	5	3,7	3,2	0	20	258					

Tabel B9.5: Leestijd per groep op AVI-toetskaarten gemeten in maart 1993 (gemiddelde, standaardafwijking, minimum en maximum)

toetskaart	groep 4					groep 5					groep 6				
	Gemid- delde	Standaard- afwijking	min	max	N	Gemid- delde	Standaard- afwijking	min	max	N	Gemid- delde	Standaard- afwijking	min	max	N
1A	71,8	22,7	36	154	177	63,2	11,2	45	83	18	65,0		65	65	1
1B	82,5	25,1	39	187	182	78,9	23,2	51	150	18	76,0		76	76	1
2A	79,5	26,4	39	244	456	76,9	17,8	40	121	49	76,0	14,1	57	90	4
2B	80,4	27,5	41	257	472	80,6	24,6	44	187	48	73,0	4,2	69	78	4
3A	89,1	32,8	40	256	610	82,3	20,3	42	165	126	81,3	15,3	54	115	16
3B	97,6	34,5	46	257	624	93,1	23,3	45	189	129	92,8	18,0	69	126	16
4A	93,0	28,8	47	245	611	86,6	21,7	48	164	223	88,4	18,7	57	157	44
4B	89,1	24,4	47	207	622	86,9	18,9	52	174	225	88,0	16,5	59	139	44
5A	88,8	19,4	54	172	404	89,5	17,6	55	172	306	89,0	14,5	62	140	102
5B	86,7	19,5	48	186	418	90,5	18,2	43	164	316	91,5	16,5	64	154	107
6A	73,3	15,5	49	167	221	74,1	12,9	46	123	367	74,0	15,0	45	150	175
6B	75,4	15,4	50	181	220	78,9	14,4	51	163	378	80,8	15,2	50	135	178
7A	106,5	20,1	68	194	143	105,2	18,2	50	181	370	105,0	17,3	70	179	293
7B	98,6	17,0	63	153	133	101,1	17,0	67	187	381	103,1	17,3	62	164	301
8A	96,4	15,8	69	130	52	94,3	16,9	60	164	453	90,1	16,5	57	167	662
8B	84,1	13,4	55	122	54	82,2	14,1	54	164	471	78,3	15,0	52	220	683
9A	105,3	14,8	80	134	27	102,4	16,5	67	170	314	100,4	17,8	61	173	578
9B	105,3	14,7	77	142	28	104,8	16,0	72	221	329	102,7	17,3	60	171	596

Tabel B9.6: Aantal fouten per groep op AVI-toetskaarten gemeten in maart 1993 (gemiddelde, standaardafwijking, minimum en maximum)

toetskaart	groep 4					groep 5					groep 6				
	Gemid- delde	Standaard- afwijking	min	max	N	Gemid- delde	Standaard- afwijking	min	max	N	Gemid- delde	Standaard- afwijking	min	max	N
1A	2,6	3,0	0	16	177	2,4	2,3	0	8	18	2,0		2	2	1
1B	2,7	2,8	0	16	182	2,2	2,0	0	7	18	3,0		3	3	1
2A	2,0	2,3	0	17	456	2,0	2,0	0	12	49	4,8	3,5	1	9	4
2B	1,7	2,0	0	14	472	2,3	2,0	0	8	48	3,0	1,8	1	5	4
3A	3,1	3,0	0	18	610	2,8	2,3	0	10	127	3,3	2,8	0	10	16
3B	3,7	3,3	0	27	625	4,0	3,2	0	17	129	3,9	2,7	1	9	16
4A	3,2	2,8	0	14	612	2,8	2,3	0	12	223	3,4	3,3	0	17	44
4B	2,5	3,0	0	41	623	2,4	2,4	0	17	225	2,5	2,7	0	12	44
5A	2,2	2,2	0	15	405	2,4	2,1	0	14	308	2,3	1,9	0	7	102
5B	2,5	2,3	0	14	419	2,7	2,4	0	16	316	2,6	2,3	0	11	107
6A	1,5	1,7	0	12	222	1,7	1,7	0	12	367	1,8	1,7	0	9	175
6B	1,6	1,7	0	9	221	1,8	1,8	0	11	378	2,1	1,9	0	10	179
7A	3,0	2,8	0	13	143	2,8	2,5	0	16	372	3,2	2,7	0	24	293
7B	2,9	3,2	0	23	133	3,4	2,8	0	22	381	3,6	2,6	0	15	301
8A	2,3	2,2	0	9	53	2,8	2,9	0	22	454	2,5	2,3	0	17	663
8B	2,1	1,9	0	11	54	2,9	2,9	0	22	472	2,5	2,2	0	19	683
9A	2,5	2,5	0	11	27	2,7	2,4	0	15	315	2,8	2,5	0	14	580
9B	2,7	2,9	0	15	28	3,1	3,7	0	38	329	2,9	2,5	0	15	599

BIJLAGE 10 UITGETESTE GRENSWAARDEN VOOR FOUTEN

Om voor iedere toetskaart de onder- en bovengrens voor de fouten te bepalen zijn twee versies uitgetest (zie paragraaf 3.2.4). In onderstaande tabel staan deze versies weergegeven. De criteria geven telkens de bovengrenzen aan. Bijvoorbeeld: om nog de score voldoende te krijgen op kaart 1A mogen volgens versie 1 maximaal 5 fouten gemaakt worden. Naast het aantal fouten is tevens aangegeven welk percentage dat aantal is van het totaal aantal woorden in de tekst.

Tabel B10.1: Uitgeteste foutencriteria

kaart	gemiddelde plus 0,5 standaardafwijking				gemiddelde plus 1 standaardafwijking			
	voldoende		goed		voldoende		goed	
	fout	%	fout	%	fout	%	fout	%
1A	5	4,8	3	2,9	7	6,7	4	3,8
1B	4	3,8	3	2,9	6	5,8	4	3,8
2A	4	3,1	3	2,3	5	3,9	4	3,1
2B	3	2,4	2	1,6	4	3,2	3	2,4
3A	5	3,8	4	3,0	7	5,3	5	3,8
3B	6	4,2	5	3,5	8	5,6	6	4,2
4A	5	3,4	5	3,4	7	4,7	6	4,0
4B	4	2,6	3	1,9	5	3,2	5	3,2
5A	4	2,4	3	1,8	5	3,0	4	2,4
5B	4	2,6	3	1,9	5	3,2	4	2,6
6A	3	2,0	2	1,3	4	2,7	3	2,0
6B	2	1,3	2	1,3	3	2,0	3	2,0
7A	5	2,4	4	1,9	6	2,9	5	2,4
7B	5	2,5	4	2,0	6	3,0	6	3,0
8A	5	2,5	4	2,0	7	3,4	5	2,5
8B	4	2,2	3	1,6	5	2,7	4	2,2
9A	4	1,7	3	1,3	6	2,6	4	1,7
9B	4	1,8	3	1,3	6	2,7	4	1,8

BIJLAGE 11 GEBRUIKTE TOETSEN VOOR TECHNISCH LEZEN, BEGRIJPEND LEZEN EN WOORDENSCHAT

toets	afnametijdstip	scorebereik	
		minimum	maximum
TECHNISCH LEZEN			
Een-Minuut-Test vorm A (Brus & Voeten)	voorgaer 3,4,5,6 najaar 4,5,6	0	116
Een-Minuut-Test vorm B (Brus & Voeten)	voorgaer 3,4,5,6 najaar 4,5,6	0	116
AVI-toetskaarten (KPC)	voorgaer 3,4,5,6 najaar 4,5,6	<1	>9
BEGRIJPEND LEZEN			
Lees en Begrijp 1 (Cito)	voorgaer groep 3	0	27
Lees en Begrijp 2 (Cito)	voorgaer groep 3 najaar groep 4	0	30
Begrijpend Leestest 4 (Aarmoutse)	najaar groep 4 voorgaer groep 4	0	36
Begrijpend Lezen M3 (Cito)	voorgaer groep 4 najaar groep 5	0	25
Begrijpend Leestest 5 (Aarmoutse)	najaar groep 5 voorgaer groep 5	0	30
Begrijpend Lezen M4 (Cito)	voorgaer groep 5 najaar groep 6	0	25
Begrijpend Leestest 6 (Aarmoutse)	najaar groep 6 voorgaer groep 6	0	30
Begrijpend Lezen E4 (Cito)	voorgaer groep 6	0	25
WOORDENSCHAT			
Begrijpendtest Woorden (Aarmoutse)	voorgaer groep 3	0	40
Woordenschattest 3 (Aarmoutse)	voorgaer groep 3 najaar groep 4	0	26
Woordenschattest 4 (Aarmoutse)	najaar groep 4 voorgaer groep 4	0	35
Woordenschattest 5 (Aarmoutse)	voorgaer groep 4 najaar groep 5	0	36
Woordenschattest 3 (Stijnen)	najaar groep 5 voorgaer groep 5	0	55
Woordenschattest 6 (Aarmoutse)	voorgaer groep 5 najaar groep 6	0	33
Woordenschattest 4 (Stijnen)	najaar groep 6 voorgaer groep 6	0	61
Woordenschattest 5 (Stijnen)	voorgaer groep 6	0	61

BIJLAGE 12 DESCRIPTIEVE STATISTIEKEN VAN DE TOETSEN VOOR TECHNISCH LEZEN, BEGRIJPEND LEZEN EN WOORDENSCHAT IN HET OORSPRONKELIJKE BESTAND

Tabel B12.1: Descriptieve statistieken van de toetsen voor technisch lezen per cohort per tijdstip in het oorspronkelijk bestand

	cohort 3			cohort 4			cohort 5		
	Gemid- delde	Standaard- afwijking	N	Gemid- delde	Standaard- afwijking	N	Gemid- delde	Standaard- afwijking	N
MAART GROEP 3									
EMT versie A	25,1	12,1	857						
EMT versie B	23,5	11,8	860						
AVI versie A	1,1	1,4	1038						
AVI versie B	1,1	1,4	1012						
OKTOBER GROEP 4									
EMT versie A	31,6	13,5	954						
EMT versie B	30,5	13,9	953						
AVI versie A	3,0	1,9	950						
MAART GROEP 4									
EMT versie A	43,0	13,9	953	44,8	14,5	911			
EMT versie B	41,9	14,3	953	43,2	14,7	913			
AVI versie A	4,5	2,1	905	4,8	2,3	978			
AVI versie B	4,5	2,1	928	5,0	2,4	973			
OKTOBER GROEP 5									
EMT versie A				52,4	14,8	850			
EMT versie B				51,1	14,8	852			
AVI versie A				5,9	2,3	847			
MAART GROEP 5									
EMT A				59,1	13,8	850	58,8	13,6	848
EMT B				58,4	14,2	849	57,9	13,5	835
AVI versie A				7,0	2,1	806	6,9	1,9	917
AVI versie B				6,9	2,2	830	7,0	2,0	917
OKTOBER GROEP 6									
EMT versie A							63,8	13,6	810
EMT versie B							62,7	13,7	809
AVI versie A							7,6	1,9	805
MAART GROEP 6									
EMT versie A							69,1	13,4	807
EMT versie B							68,3	13,6	807
AVI versie A							8,3	1,7	770
AVI versie B							8,2	1,8	795

Tabel B12.2: Descriptieve statistieken van de toetsen voor begrijpend lezen per cohort per tijdstip in het oorspronkelijk bestand

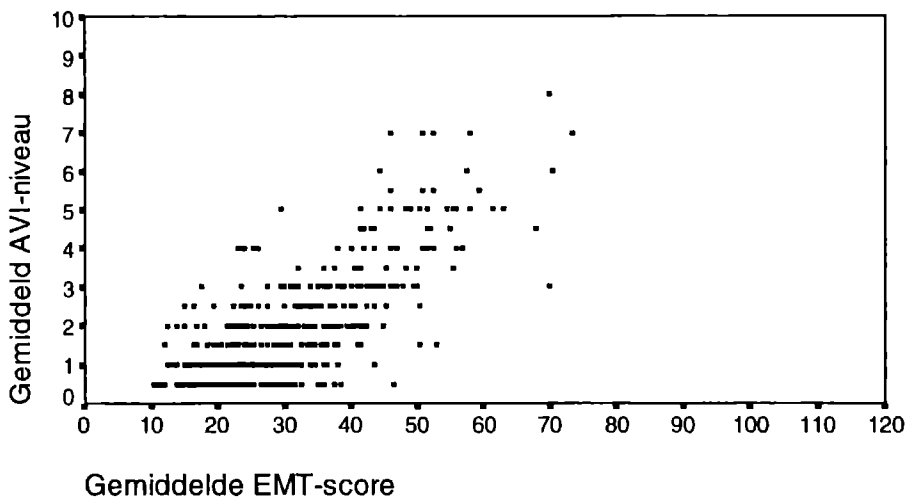
	cohort 3			cohort 4			cohort 5		
	Gemid- delde	Standaard- afwijking	N	Gemid- delde	Standaard- afwijking	N	Gemid- delde	Standaard- afwijking	N
MAART GROEP 3									
Lees en Begrijp 1 (Cito)	22,7	4,0	861						
Lees en Begrijp 2 (Cito)	21,2	7,5	815						
OKTOBER GROEP 4									
Lees en Begrijp 2 (Cito)	22,4	6,9	930						
Begrijpend Leestest 4 (Aamoutse)	21,2	7,5	928						
MAART GROEP 4									
Begrijpend Leestest 4 (Aamoutse)	26,8	6,2	919	27,2	6,0	730			
Begrijpend Lezen M3 (Cito)	16,5	5,4	911	16,9	5,1	721			
OKTOBER GROEP 5									
Begrijpend Lezen M3 (Cito)				18,4	5,4	835			
Begrijpend Leestest 5 (Aamoutse)				18,7	6,5	798			
MAART GROEP 5									
Begrijpend Leestest 5 (Aamoutse)				22,1	5,7	814	22,6	5,3	719
Begrijpend Lezen M4 (Cito)				17,3	5,2	826	17,7	4,6	720
OKTOBER GROEP 6									
Begrijpend Lezen M4 (Cito)							19,3	4,6	797
Begrijpend Leestest 6 (Aamoutse)							20,0	5,4	796
MAART GROEP 6									
Begrijpend Leestest 6 (Aamoutse)							22,9	5,2	790
Begrijpend Lezen E4 (Cito)							17,6	4,3	788

Tabel B12.3: Descriptieve statistieken van de toetsen voor woordenschat per cohort per tijdstip oorspronkelijk bestand

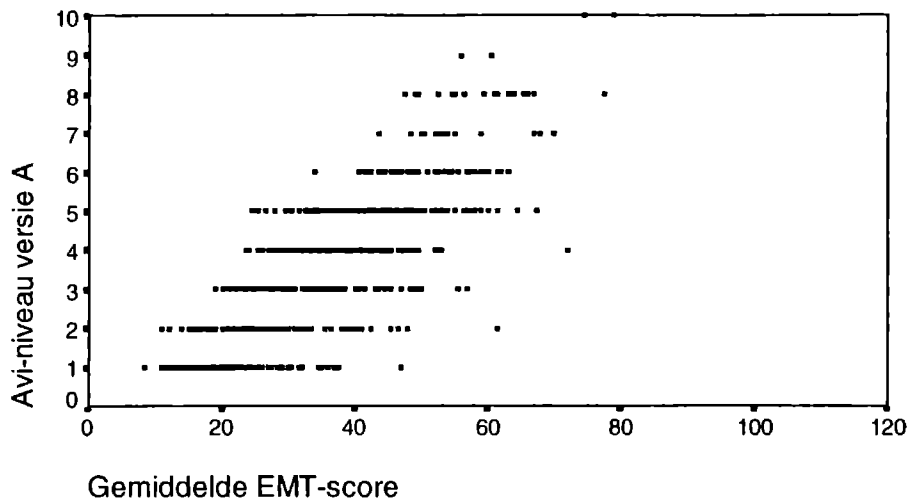
	cohort 3			cohort 4			cohort 5		
	Gemid- delde	Standaard- afwijking	N	Gemid- delde	Standaard- afwijking	N	Gemid- delde	Standaard- afwijking	N
MAART GROEP 3									
Begnp test Woorden (Aarmoutse)	28,7	6,1	816						
Woordenschattest 3 (Aarmoutse)	19,4	4,8	855						
OKTOBER GROEP 4									
Woordenschattest 3 (Aarmoutse)	21,4	4,3	927						
Woordenschattest 4 (Aarmoutse)	21,6	7,3	930						
MAART GROEP 4									
Woordenschattest 4 (Aarmoutse)	27,1	6,0	924	27,5	5,8	799			
Woordenschattest 5 (Aarmoutse)	20,9	6,8	918	20,8	7,0	801			
OKTOBER GROEP 5									
Woordenschattest 5 (Aarmoutse)				23,8	6,9	832			
Woordenschattest 3 (Stijnen)				38,8	8,6	830			
MAART GROEP 5									
Woordenschattest 3 (Stijnen)				43,7	7,4	825	43,4	8,2	765
Woordenschattest 6 (Aarmoutse)				20,1	5,1	820	19,4	5,2	735
OKTOBER GROEP 6									
Woordenschattest 6 (Aarmoutse)							21,4	5,6	773
Woordenschattest 4 (Stijnen)							40,0	9,1	789
MAART GROEP 6									
Woordenschattest 4 (Stijnen)							44,5	8,0	757
Woordenschattest 5 (Stijnen)							39,0	8,4	747

BIJLAGE 13 SPREIDINGSDIAGRAMMEN VAN SCORES OP DE AVI-TOETS- KAARTEN MET SCORES OP DE EMT

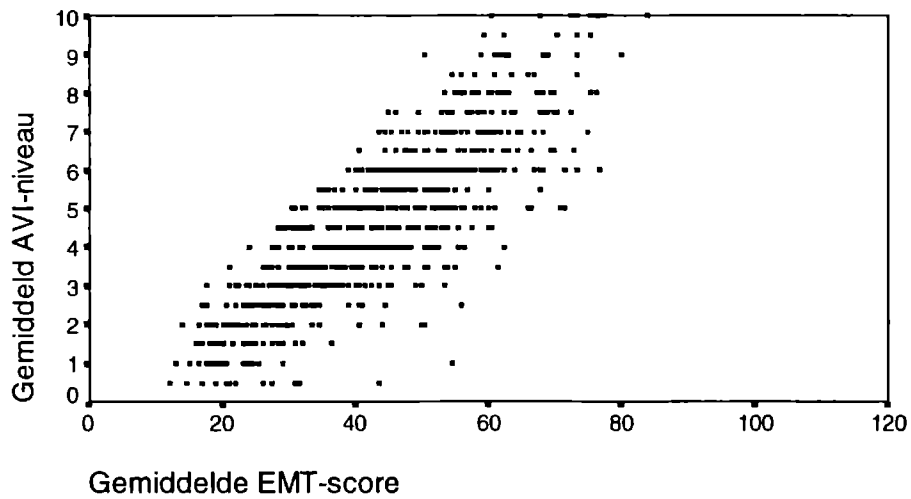
In deze bijlage zijn per cohort per periode de spreidingsdiagrammen weergegeven van de scores op de AVI-toetskaarten met de scores op de Eén-Minuut-Test. De scores op de AVI-toetskaarten betreffen voor het voorjaar van 1992 en 1993 de gemiddelde AVI-niveaus van de versies A en B; voor het najaar van 1992 betreft het het AVI-niveau van versie A. De weergegeven EMT-scores zijn telkens de gemiddelden van versies A en B.



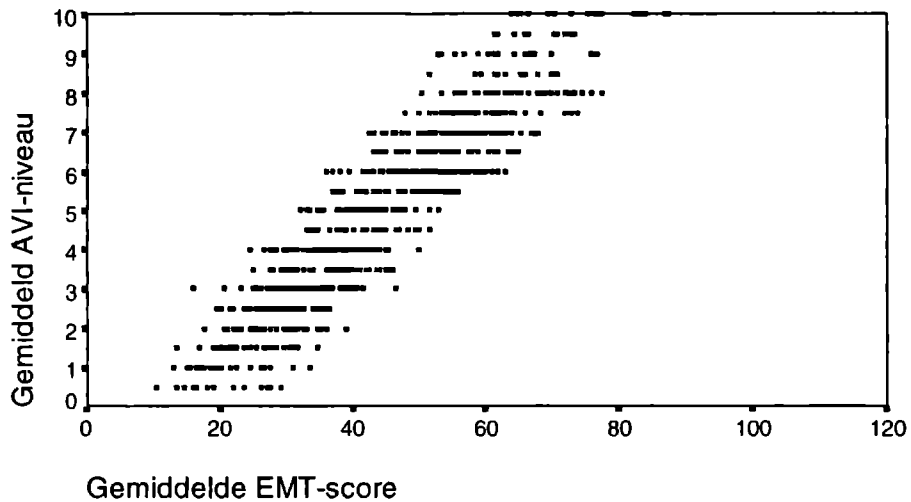
Figuur 1: Spreidingsdiagram voor cohort 3 voorjaar 1992



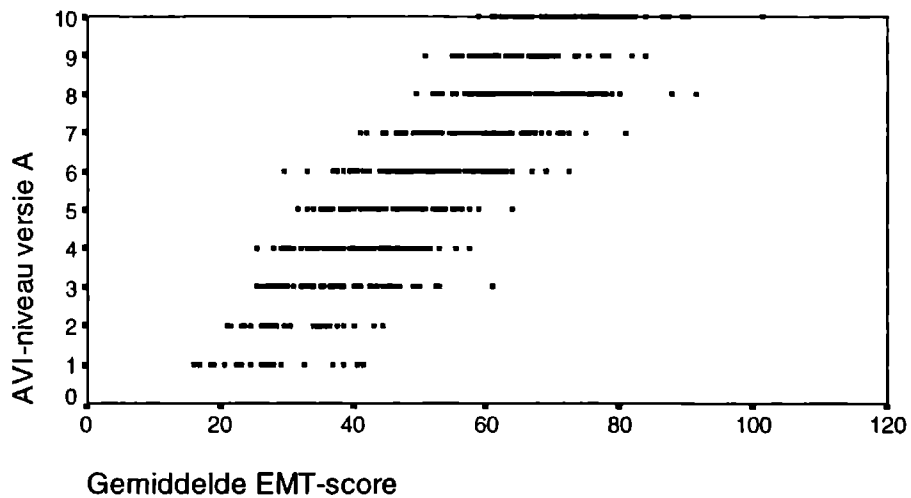
Figuur 2: Spreidingsdiagram voor cohort 3 najaar 1992



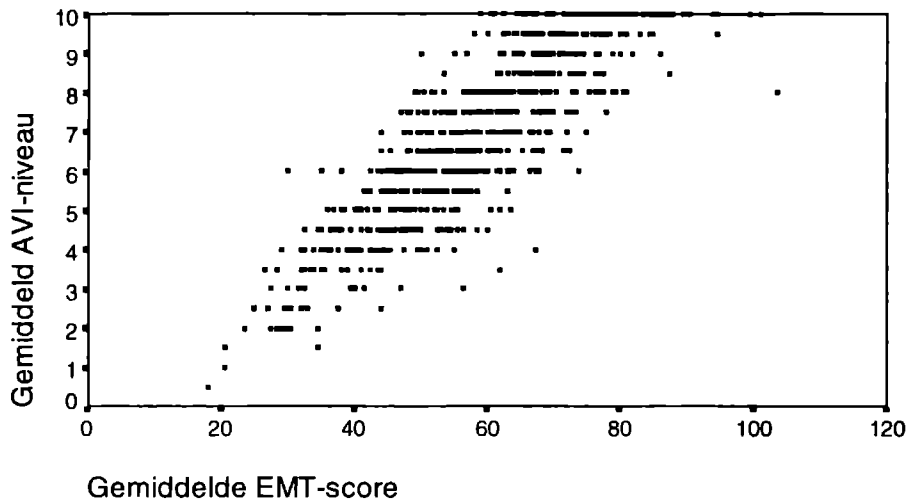
Figuur 3: Spreidingsdiagram voor cohort 3 voorjaar 1993



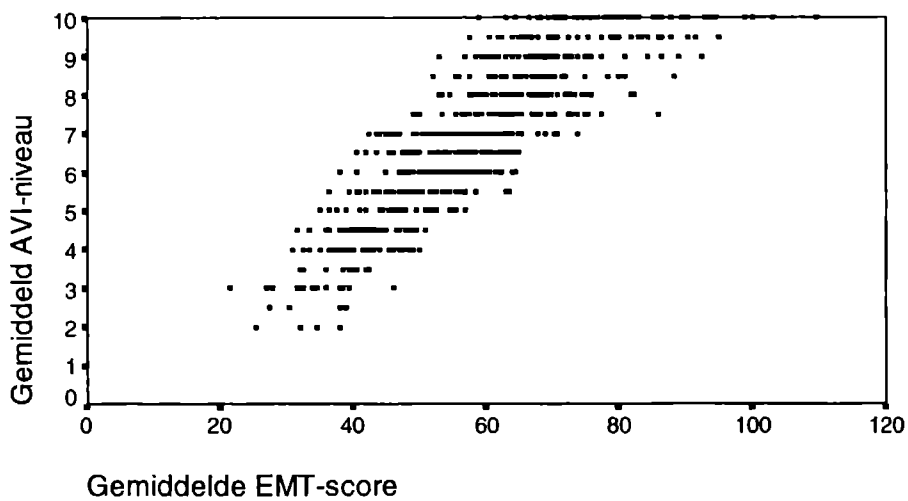
Figuur 4: Spreidingsdiagram voor cohort 4 voorjaar 1992



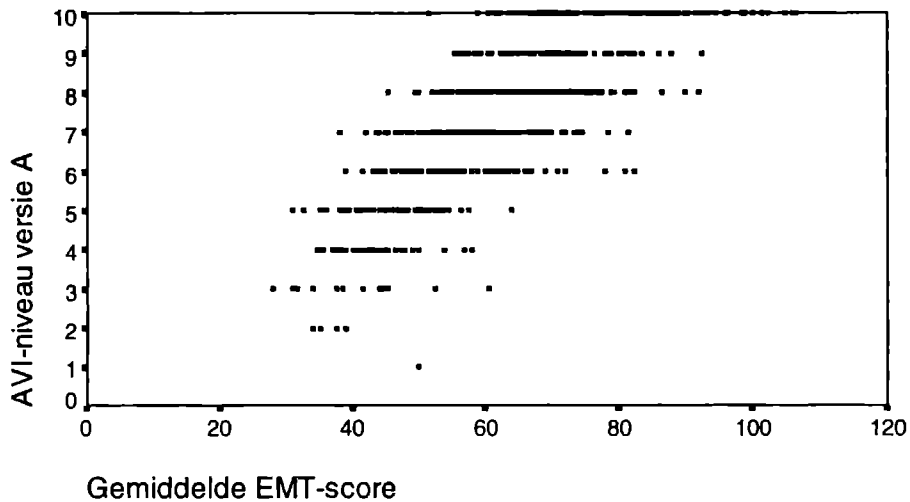
Figuur 5: Spreidingsdiagram voor cohort 4 najaar 1992



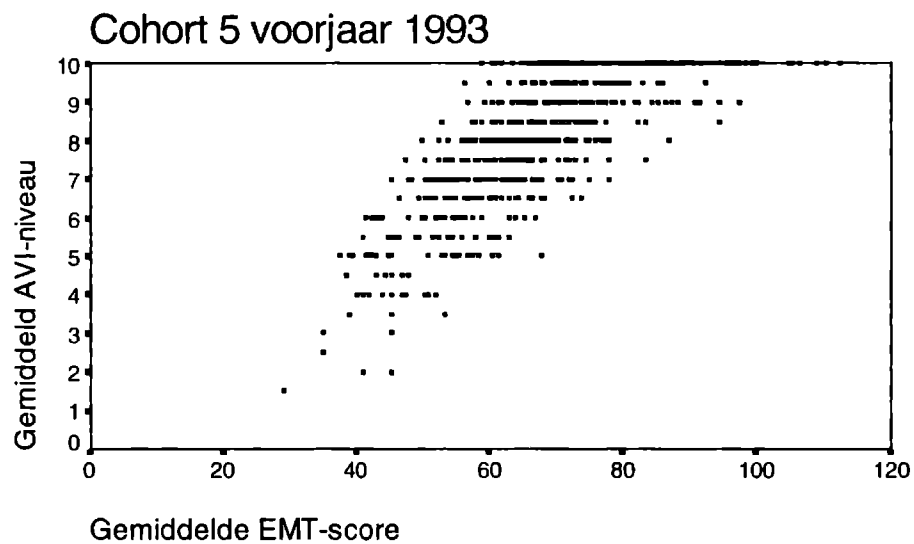
Figuur 6: Spreidingsdiagram voor cohort 4 voorjaar 1993



Figuur 7: Spreidingsdiagram voor cohort 5 voorjaar 1992



Figuur 8: Spreidingsdiagram voor cohort 5 najaar 1992



Figuur 9: Spreidingsdiagram voor cohort 5 voorjaar 1993

BIJLAGE 14 ONGESTANDAARDISEERDE PARAMETERSCHATTINGEN IN HET UITEINDELIJKE MODEL VOOR PARALLELLITEIT AVI EN EMT

Deze bijlage bevat de ongestandaardiseerde parameterschattingen voor model A2 per cohort per periode. Voor een beschrijving van het model verwijzen we naar tabel 4.8 in paragraaf 4.2.2.

Tabel B14.1: Factorloadingen van de geobserveerde variabelen op de latente factoren per cohort per periode in model A2 (Lambda-matrix)

		voorjaar 1992		najaar 1992		voorjaar 1993	
		EMT	AVI	EMT	AVI	EMT	AVI
COHORT 3							
Emt versie A	Parameter	11,716 ^a	-	13,459 ^c	-	13,829 ^d	-
	(Se)	(0,275)		(0,315)		(0,325)	
	t-waarde	42,669		42,761		42,561	
Emt versie B	Parameter	11,716 ^a	-	13,459 ^c	-	13,829 ^d	-
	(Se)	(0,275)		(0,315)		(0,325)	
	t-waarde	42,669		42,761		42,561	
AVI versie A	Parameter	-	1,351 ^b	-	1,949	-	1,953 ^e
	(Se)		(0,033)		(0,045)		(0,049)
	t-waarde		41,130		43,543		39,942
AVI versie B	Parameter	-	1,351 ^b			-	1,953 ^e
	(Se)		(0,033)				(0,049)
	t-waarde		41,130				39,942
COHORT 4							
Emt versie A	Parameter	14,325 ^f	-	14,358 ^h	-	13,595 ⁱ	-
	(Se)	(0,339)		(0,360)		(0,341)	
	t-waarde	42,259		39,883		39,832	
Emt versie B	Parameter	14,325 ^f	-	14,358 ^h	-	13,595 ⁱ	-
	(Se)	(0,339)		(0,360)		(0,341)	
	t-waarde	42,259		39,883		39,832	
AVI versie A	Parameter	-	2,225 ^g	-	2,320	-	1,994 ^j
	(Se)		(0,054)		(0,056)		(0,053)
	t-waarde		41,046		41,110		37,613
AVI versie B	Parameter	-	2,225 ^g			-	1,994 ^j
	(Se)		(0,054)				(0,053)
	t-waarde		41,046				37,613

		voorgaar 1992		najaar 1992		voorgaar 1993	
		EMT	AVI	EMT	AVI	EMT	AVI
COHORT 5							
Emt versie A	Parameter	12,979 ^k	-	13,128 ^m	-	13,101 ⁿ	-
	(Se)	(0,325)		(0,340)		(0,339)	
	t-waarde	39,873		38,598		38,366	
Emt versie B	Parameter	12,979 ^k	-	13,128 ^m	-	13,101 ⁿ	-
	(Se)	(0,325)		(0,340)		(0,339)	
	t-waarde	39,873		38,598		38,366	
AVI versie A	Parameter	-	1,806 ^d	-	1,910	-	1,641 ^o
	(Se)		(0,047)		(0,048)		(0,045)
	t-waarde		38,533		40,125		36,834
AVI versie B	Parameter	-	1,806 ^d			-	1,641 ^o
	(Se)		(0,047)				(0,045)
	t-waarde		38,533				36,834

N.B.: De letters a t/m o geven aan welke parameters aan elkaar gelijkgesteld zijn

Tabel B14.2: Varianties en covarianties van de residuen van de latente factoren per cohort per periode in model A2 (Psi-matrix)

		voorgaar 1992		najaar 1992		voorgaar 1993	
		EMT	AVI	EMT	AVI	EMT	AVI
COHORT 3							
EMT	Parameter	1,000		1,000		1,000	
	(Se)						
	t-waarde						
AVI	Parameter	0,855	1,000	0,861	1,000	0,892	1,000
	(Se)	(0,010)		(0,009)		(0,009)	
	t-waarde	84,141		98,583		103,198	
COHORT 4							
EMT	Parameter	1,000		1,000		1,000	
	(Se)						
	t-waarde						
AVI	Parameter	0,956	1,000	0,875	1,000	0,906	1,000
	(Se)	(0,004)		(0,009)		(0,009)	
	t-waarde	216,709		101,806		105,089	
COHORT 5							
EMT	Parameter	1,000		1,000		1,000	
	(Se)						
	t-waarde						
AVI	Parameter	0,899	1,000	0,807	1,000	0,830	1,000
	(Se)	(0,009)		(0,013)		(0,013)	
	t-waarde	103,870		61,883		62,461	

Tabel B14.3: Error-varianties (Theta-epsilon matrix) en verklaarde varianties (R^2) van de geobserveerde variabelen per cohort per periode in model A2

		maart 1992		oktober 1992		maart 1993	
		Theta-epsil- lon	R ²	Theta-epsil- lon	R ²	Theta-epsil- lon	R ²
COHORT 3							
Emit versie A	Parameter	6,291 ^a	0,956	6,567 ^c	0,965	7,713 ^d	0,961
	(Se)	(0,288)		(0,302)		(0,355)	
	t-waarde	21,829		21,772		21,714	
Emit versie B	Parameter	6,291 ^a	0,956	6,567 ^c	0,965	7,713 ^d	0,961
	(Se)	(0,288)		(0,302)		(0,355)	
	t-waarde	21,829		21,772		21,714	
AVI versie A	Parameter	0,218 ^b	0,893	-	1,000	0,641 ^e	0,856
	(Se)	(0,010)		(0,030)			
	t-waarde	21,829		21,714			
AVI versie B	Parameter	0,218 ^b	0,893			0,641 ^e	0,856
	(Se)	(0,010)		(0,030)			
	t-waarde	21,829		21,714			
COHORT 4							
Emit versie A	Parameter	9,670 ^f	0,955	12,493 ^h	0,943	11,018 ⁱ	0,944
	(Se)	(0,447)		(0,608)		(0,537)	
	t-waarde	21,633		20,555		20,518	
Emit versie B	Parameter	9,670 ^f	0,955	12,493 ^h	0,943	11,018 ⁱ	0,944
	(Se)	(0,447)		(0,608)		(0,537)	
	t-waarde	21,633		20,555		20,518	
AVI versie A	Parameter	0,523 ^g	0,905	-	1,000	0,696 ^j	0,851
	(Se)	(0,024)		(0,034)			
	t-waarde	21,633		20,518			
AVI versie B	Parameter	0,523 ^g	0,905			0,696 ^j	0,851
	(Se)	(0,024)		(0,034)			
	t-waarde	21,633		20,518			
COHORT 5							
Emit versie A	Parameter	11,488 ^k	0,936	13,385 ^m	0,928	11,864 ⁿ	0,935
	(Se)	(0,557)		(0,667)		(0,593)	
	t-waarde	20,628		20,062		20,012	
Emit versie B	Parameter	11,488 ^k	0,936	13,385 ^m	0,928	11,864 ⁿ	0,935
	(Se)	(0,557)		(0,667)		(0,593)	
	t-waarde	20,628		20,062		20,012	
AVI versie A	Parameter	0,446 ^l	0,880	-	1,000	0,449 ^o	0,857
	(Se)	(0,022)		(0,022)			
	t-waarde	20,628		20,012			
AVI versie B	Parameter	0,446 ^l	0,880			0,449 ^o	0,857
	(Se)	(0,022)		(0,022)			
	t-waarde	20,628		20,012			

N.B.. De letters a t/m o geven aan welke parameters aan elkaar gelijkgesteld zijn

BIJLAGE 15 ONGESTANDAARDISEERDE PARAMETERSCHATTINGEN IN HET UITEINDELIJKE MODEL VOOR PARALLELLITEIT AVI EN EMT BIJ SIMULTANE ANALYSES VOOR TWEE COHORTEN

In deze bijlage zijn de ongestandaardiseerde parameterschattingen weergegeven uit de simultane Lisrel-analyses. De tabellen B15.1 tot en met B15.5 betreffen de schattingen in model A10b voor het voorjaar van groep 4 in de cohorten 3 en 4. De tabellen B15.6 tot en met B15.10 betreffen de schattingen uit model A10b voor het voorjaar van groep 5 in de cohorten 4 en 5. Voor een beschrijving van deze modellen verwijzen we naar tabel 4.9 in paragraaf 4.2.2.

Tabel B15.1: Gefixeerde factorladingen van de geobserveerde variabelen op de latente factoren (Lambda-matrix) voor het voorjaar van groep 4 in model A10b

		cohort 3		cohort 4	
		EMT	AVI	EMT	AVI
Emt versie A	Parameter	1,000	-	1,000	-
Emt versie B	Parameter	1,000	-	1,000	-
AVI versie A	Parameter	-	1,000	-	1,000
AVI versie B	Parameter	-	1,000	-	1,000

Tabel B15.2: Varianties en covarianties van de residuen van de latente factoren (Psi-matrix) voor het voorjaar van groep 4 in model A10b

		cohort 3		cohort 4	
		EMT	AVI	EMT	AVI
EMT	Parameter	191,888		205,949	
	(Se)	(9,017)		(9,747)	
	t-waarde	21,281		21,129	
AVI	Parameter	24,174	3,823	30,590	4,965
	(Se)	(1,216)	(0,191)	(1,476)	(0,242)
	t-waarde	19,872	19,975	20,730	20,512

Tabel B15.3: Error-varianties (Theta-epsilon matrix) en verklaarde varianties (R^2) van de geobserveerde variabelen voor het voorjaar van groep 4 in model A10b

		cohort 3		cohort 4	
		Theta-epsilon	R^2	Theta-epsilon	R^2
Emt versie A	Parameter	7,740 ^a	0,961	9,714 ^c	0,955
	(Se)	(0,356)		(0,449)	
	t-waarde	21,714		21,633	
Emt versie B	Parameter	7,740 ^a	0,961	9,714 ^c	0,955
	(Se)	(0,356)		(0,449)	
	t-waarde	21,714		21,633	
AVI versie A	Parameter	0,641 ^b	0,856	0,529 ^d	0,904
	(Se)	(0,030)		(0,024)	
	t-waarde	21,714		21,633	
AVI versie B	Parameter	0,641 ^b	0,856	0,529 ^d	0,904
	(Se)	(0,030)		(0,024)	
	t-waarde	21,714		21,633	

N B De letters a tot en met d geven aan welke parameters aan elkaar gelijkgesteld zijn

Tabel B15.4: Gemiddelden op de geobserveerde variabelen (Tau-Y-matrix) voor het voorjaar van groep 4 in model A10b

		cohort 3	cohort 4
Emt versie A	Parameter	43,945 ^a	43,945 ^a
	(Se)	(0,332)	(0,332)
	t-waarde	132,309	132,309
Emt versie B	Parameter	42,695 ^b	42,695 ^b
	(Se)	(0,332)	(0,332)
	t-waarde	128,545	128,545
AVI versie A	Parameter	4,618 ^c	4,618 ^c
	(Se)	(0,054)	(0,054)
	t-waarde	86,119	86,119
AVI versie B	Parameter	4,618 ^c	4,618 ^c
	(Se)	(0,054)	(0,054)
	t-waarde	86,119	86,119

N B De letters a tot en met c geven aan welke parameters aan elkaar gelijkgesteld zijn

Tabel B15.5: Verschillen in gemiddelden op de latente variabelen van cohort 4 ten opzichte van cohort 3 (Alpha-matrix) voor het voorjaar van groep 4 in model A10b

		Emt	Avi
cohort 4	Parameter	-	0,197
	(Se)		(0,046)
	t-waarde		4,281

Tabel B15.6: Gefixeerde factorladingen van de geobserveerde variabelen op de latente factoren (Lambda-matrix) voor het voorjaar van groep 5 in model A10b

		cohort 4		cohort 5	
		EMT	AVI	EMT	AVI
Emt versie A	Parameter	1,000	-	1,000	-
Emt versie B	Parameter	1,000	-	1,000	-
AVI versie A	Parameter	-	1,000	-	1,000
AVI versie B	Parameter	-	1,000	-	1,000

Tabel B15.7: Varianties en covarianties van de residuen van de latente factoren (Psi-matrix) voor het voorjaar van groep 5 in model A10b

		cohort 4		cohort 5	
		EMT	AVI	EMT	AVI
EMT	Parameter	184,848		168,469	
	(Se)	(9,282)		(8,451)	
	t-waarde	19,915		19,936	
AVI	Parameter	24,573	3,974	21,067	3,261
	(Se)	(1,302)	(0,212)	(1,111)	(0,169)
	t-waarde	18,873	18,789	18,958	19,264

Tabel B15.8: Error-varianties (Theta-epsilon-matrix) en verklaarde varianties (R^2) van de geobserveerde variabelen voor het voorjaar van groep 5 in model A10b

		cohort 4		cohort 5	
		Theta-epsilon	R^2	Theta-epsilon	R^2
Emt versie A	Parameter	11,032 ^a	0,944	11,502 ^c	0,936
	(Se)	(0,538)		(0,558)	
	t-waarde	20,518		20,628	
Emt versie B	Parameter	11,032 ^a	0,944	11,502 ^c	0,936
	(Se)	(0,538)		(0,558)	
	t-waarde	20,518		20,628	
AVI versie A	Parameter	0,703 ^b	0,850	0,447 ^d	0,879
	(Se)	(0,034)		(0,022)	
	t-waarde	20,518		20,628	
AVI versie B	Parameter	0,703 ^b	0,850	0,447 ^d	0,879
	(Se)	(0,034)		(0,022)	
	t-waarde	20,518		20,628	

N.B: De letters a tot en met d geven aan welke parameters aan elkaar gelijkgesteld zijn

Tabel B15.9: Gemiddelden op de geobserveerde variabelen (Tau-Y-matrix) voor het voorjaar van groep 5 in model A10b

		cohort 4	cohort 5
Emt versie A	Parameter	59,033 ^a	59,033 ^a
	(Se)	(0,332)	(0,332)
	t-waarde	177,624	177,624
Emt versie B	Parameter	58,198 ^b	58,198 ^b
	(Se)	(0,332)	(0,332)
	t-waarde	175,112	175,112
AVI versie A	Parameter	6,950 ^c	6,950 ^c
	(Se)	(0,048)	(0,048)
	t-waarde	145,795	145,795
AVI versie B	Parameter	6,950 ^c	6,950 ^c
	(Se)	(0,048)	(0,048)
	t-waarde	145,795	145,795

N.B.. De letters a tot en met c geven aan welke parameters aan elkaar gelijkgesteld zijn

BIJLAGE 16 ONGESTANDAARDISEERDE PARAMETERSCHATTINGEN IN HET UITEINDELIJKE MODEL PER PERIODE VOOR DE RELATIE VAN AVI MET BEGRIJPEND LEZEN EN WOORDENSCHAT

Deze bijlage bevat de ongestandaardiseerde parameterschattingen voor model B per cohort per periode. Voor een beschrijving van het model verwijzen we naar tabel 4.11 in paragraaf 4.3.2.

Tabel B16.1: Factorladingen van de geobserveerde variabelen op de latente factoren per cohort per periode in model B (Lambda-matrix)

		EMT	AVI	BL	WS
COHORT 3 VOORJAAR 1992					
Emt versie A	Parameter	11,716 ^a			
	(Se)	(0,275)			
	t-waarde	42,669			
Emt versie B	Parameter	11,716 ^a			
	(Se)	(0,275)			
	t-waarde	42,669			
AVI versie A	Parameter		1,351 ^b		
	(Se)		(0,033)		
	t-waarde		41,130		
AVI versie B	Parameter		1,351 ^b		
	(Se)		(0,033)		
	t-waarde		41,130		
Lees en Begrijp 1 (Cito)	Parameter			3,098	
	(Se)			(0,113)	
	t-waarde			27,461	
Lees en Begrijp 2 (Cito)	Parameter			6,137	
	(Se)			(0,207)	
	t-waarde			29,654	
Begrippentest Woorden (Aarmoutse)	Parameter				4,747
	(Se)				(0,175)
	t-waarde				27,161
Woordenschattest 3 (Aarmoutse)	Parameter				4,032
	(Se)				(0,134)
	t-waarde				30,080

		EMT	AVI	BL	WS
COHORT 3 NAJAAR 1992					
Emt versie A	Parameter	13,459 ^c			
	(Se)	(0,315)			
	t-waarde	42,761			
Emt versie B	Parameter	13,459 ^c			
	(Se)	(0,315)			
	t-waarde	42,761			
AVI versie A	Parameter		1,949		
	(Se)		(0,045)		
	t-waarde		43,543		
Lees en Begrijp 2 (Cito)	Parameter			5,711	
	(Se)			(0,185)	
	t-waarde			30,190	
Begrijpend Leestest 4 (Aarmoutse)	Parameter			6,626	
	(Se)			(0,197)	
	t-waarde			33,644	
Woordenschattest 3 (Aarmoutse)	Parameter				3,444
	(Se)				(0,117)
	t-waarde				29,338
Woordenschattest 4 (Aarmoutse)	Parameter				6,373
	(Se)				(0,195)
	t-waarde				32,725
COHORT 3 VOORJAAR 1993					
Emt versie A	Parameter	13,829 ^d			
	(Se)	(0,325)			
	t-waarde	42,561			
Emt versie B	Parameter	13,829 ^d			
	(Se)	(0,325)			
	t-waarde	42,561			
AVI versie A	Parameter		1,953 ^e		
	(Se)		(0,049)		
	t-waarde		39,942		
AVI versie B	Parameter		1,953 ^e		
	(Se)		(0,049)		
	t-waarde		39,942		
Begrijpend Lezen M3 (Cito)	Parameter			4,209	
	(Se)			(0,154)	
	t-waarde			27,351	
Begrijpend Leestest 4 (Aarmoutse)	Parameter			5,306	
	(Se)			(0,174)	
	t-waarde			30,463	
Woordenschattest 4 (Aarmoutse)	Parameter				5,406
	(Se)				(0,160)
	t-waarde				33,857
Woordenschattest 5 (Aarmoutse)	Parameter				5,635
	(Se)				(0,188)
	t-waarde				30,029

		EMT	AVI	BL	WS
COHORT 4 VOORJAAR 1992					
Emt versie A	Parameter	14,325 ^f			
	(Se)	(0,339)			
	t-waarde	42,259			
Emt versie B	Parameter	14,325 ^f			
	(Se)	(0,339)			
	t-waarde	42,259			
AVI versie A	Parameter		2,225 ^g		
	(Se)		(0,054)		
	t-waarde		41,046		
AVI versie B	Parameter		2,225 ^g		
	(Se)		(0,054)		
	t-waarde		41,046		
Begrijpend Leestest 4 (Aarmoutse)	Parameter			4,936	
	(Se)			(0,175)	
	t-waarde			28,161	
Begrijpend Lezen M3 (Cito)	Parameter			4,088	
	(Se)			(0,146)	
	t-waarde			28,089	
Woordenschattest 4 (Aarmoutse)	Parameter				5,191
	(Se)				(0,162)
	t-waarde				31,961
Woordenschattest 5 (Aarmoutse)	Parameter				5,937
	(Se)				(0,191)
	t-waarde				31,043
COHORT 4 NAJAAR 1992					
Emt versie A	Parameter	14,358 ^h			
	(Se)	(0,360)			
	t-waarde	39,883			
Emt versie B	Parameter	14,358 ^h			
	(Se)	(0,360)			
	t-waarde	39,883			
AVI versie A	Parameter		2,320		
	(Se)		(0,056)		
	t-waarde		41,110		
Begrijpend Lezen M3 (Cito)	Parameter			4,280	
	(Se)			(0,160)	
	t-waarde			26,669	
Begrijpend Leestest 5 (Aarmoutse)	Parameter			5,792	
	(Se)			(0,188)	
	t-waarde			30,729	
Woordenschattest 5 (Aarmoutse)	Parameter				5,835
	(Se)				(0,199)
	t-waarde				29,302
Woordenschattest 3 (Stijnen)	Parameter				7,736
	(Se)				(0,242)
	t-waarde				31,940

		EMT	AVI	BL	WS
COHORT 4 VOORJAAR 1993					
Emt versie A	Parameter	13,595'			
	(Se)	(0,341)			
	t-waarde	39,832			
Emt versie B	Parameter	13,595'			
	(Se)	(0,341)			
	t-waarde	39,832			
AVI versie A	Parameter		1,994'		
	(Se)		(0,053)		
	t-waarde		37,613		
AVI versie B	Parameter		1,994'		
	(Se)		(0,053)		
	t-waarde		37,613		
Begrijpend Leestest 5 (Aarmoutse)	Parameter			4,839	
	(Se)			(0,167)	
	t-waarde			29,017	
Begrijpend Lezen M4 (Cito)	Parameter			4,419	
	(Se)			(0,153)	
	t-waarde			28,902	
Woordenschattest 3 (Stijnen)	Parameter				6,626
	(Se)				(0,227)
	t-waarde				29,162
Woordenschattest 6 (Aarmoutse)	Parameter				3,806
	(Se)				(0,161)
	t-waarde				23,706
COHORT 5 VOORJAAR 1992					
Emt versie A	Parameter	12,979 ^k			
	(Se)	(0,325)			
	t-waarde	39,873			
Emt versie B	Parameter	12,979 ^k			
	(Se)	(0,325)			
	t-waarde	39,873			
AVI versie A	Parameter		1,806 ^d		
	(Se)		(0,047)		
	t-waarde		38,533		
AVI versie B	Parameter		1,806 ^d		
	(Se)		(0,047)		
	t-waarde		38,533		
Begrijpend Leestest 5 (Aarmoutse)	Parameter			4,399	
	(Se)			(0,151)	
	t-waarde			29,165	
Begrijpend Lezen M4 (Cito)	Parameter			3,872	
	(Se)			(0,133)	
	t-waarde			29,151	
Woordenschattest 3 (Stijnen)	Parameter				5,688
	(Se)				(0,267)
	t-waarde				21,268
Woordenschattest 6 (Aarmoutse)	Parameter				3,800
	(Se)				(0,170)
	t-waarde				22,295

		EMT	AVI	BL	WS
COHORT 5 NAJAAR 1992					
Emt versie A	Parameter	13,128 ^m			
	(Se)	(0,340)			
	t-waarde	38,598			
Emt versie B	Parameter	13,128 ^m			
	(Se)	(0,340)			
	t-waarde	38,598			
AVI versie A	Parameter		1,910		
	(Se)		(0,048)		
	t-waarde		40,125		
Begrijpend Lezen M4 (Cito)	Parameter			3,769	
	(Se)			(0,142)	
	t-waarde			26,486	
Begrijpend Leestest 6 (Aamoutse)	Parameter			4,477	
	(Se)			(0,163)	
	t-waarde			27,389	
Woordenschattest 6 (Aamoutse)	Parameter				4,543
	(Se)				(0,172)
	t-waarde				26,405
Woordenschattest 4 (Stijnen)	Parameter				8,232
	(Se)				(0,267)
	t-waarde				30,831
COHORT 5 VOORJAAR 1993					
Emt versie A	Parameter	13,101 ⁿ			
	(Se)	(0,339)			
	t-waarde	38,666			
Emt versie B	Parameter	13,101 ⁿ			
	(Se)	(0,339)			
	t-waarde	38,666			
AVI versie A	Parameter		1,641 ^o		
	(Se)		(0,045)		
	t-waarde		36,834		
AVI versie B	Parameter		1,641 ^o		
	(Se)		(0,045)		
	t-waarde		36,834		
Begrijpend Leestest 6 (Aamoutse)	Parameter			4,424	
	(Se)			(0,156)	
	t-waarde			28,403	
Begrijpend Lezen E4 (Cito)	Parameter			3,557	
	(Se)			(0,129)	
	t-waarde			27,676	
Woordenschattest 4 (Stijnen)	Parameter				7,385
	(Se)				(0,224)
	t-waarde				32,909
Woordenschattest 5 (Stijnen)	Parameter				7,838
	(Se)				(0,237)
	t-waarde				33,123

N.B.: De letters a t/m o geven aan welke parameters aan elkaar gelijkgesteld zijn

Tabel B16.2: Varianties en covarianties van de residuen van de latente factoren per cohort per periode in model B (Psi-matrix)

		EMT	AVI	BL	WS
COHORT 3 VOORJAAR 1992					
EMT	Parameter	1,000			
	(Se)				
AVI	t-waarde				
	Parameter	0,855	1,000		
BL	(Se)	(0,010)			
	t-waarde	84,141			
WS	Parameter	0,630	0,560	1,000	
	(Se)	(0,024)	(0,028)		
	t-waarde	25,778	20,245		
	Parameter	0,574	0,552	0,958	1,000
	(Se)	(0,026)	(0,028)	(0,016)	
	t-waarde	21,831	19,983	60,348	
COHORT 3 NAJAAR 1992					
EMT	Parameter	1,000			
	(Se)				
AVI	t-waarde				
	Parameter	0,861	1,000		
BL	(Se)	(0,009)			
	t-waarde	98,583			
WS	Parameter	0,716	0,687	1,000	
	(Se)	(0,019)	(0,020)		
	t-waarde	37,502	34,285		
	Parameter	0,717	0,692	0,981	1,000
	(Se)	(0,020)	(0,020)	(0,011)	
	t-waarde	36,500	34,034	87,863	
COHORT 3 VOORJAAR 1993					
EMT	Parameter	1,000			
	(Se)				
AVI	t-waarde				
	Parameter	0,892	1,000		
BL	(Se)	(0,009)			
	t-waarde	103,198			
WS	Parameter	0,605	0,636	1,000	
	(Se)	(0,025)	(0,025)		
	t-waarde	24,066	25,318		
	Parameter	0,562	0,603	0,927	1,000
	(Se)	(0,025)	(0,025)	(0,014)	
	t-waarde	22,269	24,260	64,386	

		EMT	AVI	BL	WS
COHORT 4 VOORJAAR 1992					
EMT	Parameter	1,000			
	(Se)				
	t-waarde				
AVI	Parameter	0,956	1,000		
	(Se)	(0,004)			
	t-waarde	216,709			
BL	Parameter	0,611	0,667	1,000	
	(Se)	(0,025)	(0,024)		
	t-waarde	23,953	27,973		
WS	Parameter	0,521	0,562	0,905	1,000
	(Se)	(0,027)	(0,026)	(0,016)	
	t-waarde	19,322	21,584	56,983	
COHORT 4 NAJAAR 1992					
EMT	Parameter	1,000			
	(Se)				
	t-waarde				
AVI	Parameter	0,875	1,000		
	(Se)	(0,009)			
	t-waarde	101,806			
BL	Parameter	0,585	0,596	1,000	
	(Se)	(0,027)	(0,026)		
	t-waarde	21,962	23,216		
WS	Parameter	0,574	0,580	0,911	1,000
	(Se)	(0,026)	(0,025)	(0,015)	
	t-waarde	21,795	22,759	62,600	
COHORT 4 VOORJAAR 1993					
EMT	Parameter	1,000			
	(Se)				
	t-waarde				
AVI	Parameter	0,906	1,000		
	(Se)	(0,009)			
	t-waarde	105,089			
BL	Parameter	0,512	0,591	1,000	
	(Se)	(0,029)	(0,028)		
	t-waarde	17,511	21,473		
WS	Parameter	0,508	0,596	0,894	1,000
	(Se)	(0,030)	(0,028)	(0,018)	
	t-waarde	16,830	21,161	50,388	

		EMT	AVI	BL	WS
COHORT 5 VOORJAAR 1992					
EMT	Parameter (Se)	1,000			
	t-waarde				
AVI	Parameter (Se)	0,899 (0,009)	1,000		
	t-waarde	103,871			
BL	Parameter (Se)	0,438 (0,032)	0,551 (0,029)	1,000	
	t-waarde	13,850	19,349		
WS	Parameter (Se)	0,448 (0,036)	0,529 (0,034)	0,986 (0,021)	1,000
	t-waarde	12,383	15,342	47,164	
COHORT 5 NAJAAR 1992					
EMT	Parameter (Se)	1,000			
	t-waarde				
AVI	Parameter (Se)	0,807 (0,013)	1,000		
	t-waarde	61,883			
BL	Parameter (Se)	0,528 (0,030)	0,535 (0,029)	1,000	
	t-waarde	17,392	18,267		
WS	Parameter (Se)	0,519 (0,029)	0,470 (0,030)	0,891 (0,017)	1,000
	t-waarde	17,644	15,516	51,769	
COHORT 5 VOORJAAR 1993					
EMT	Parameter (Se)	1,000			
	t-waarde				
AVI	Parameter (Se)	0,830 (0,013)	1,000		
	t-waarde	62,461			
BL	Parameter (Se)	0,455 (0,032)	0,514 (0,031)	1,000	
	t-waarde	14,082	16,460		
WS	Parameter (Se)	0,479 (0,029)	0,462 (0,031)	0,890 (0,014)	1,000
	t-waarde	16,250	14,915	62,279	

Tabel B16.3: Error-varianties (Theta-epsilon matrix) en verklaarde varianties (R^2) van de geobserveerde variabelen per cohort per periode in model B

		Theta-epsilon	R ²
COHORT 3 VOORJAAR 1992			
Emt versie A	Parameter	6,291 ^a	0,956
	(Se)	(0,288)	
	t-waarde	21,829	
Emt versie B	Parameter	6,291 ^a	0,956
	(Se)	(0,288)	
	t-waarde	21,829	
AVI versie A	Parameter	0,218 ^b	0,893
	(Se)	(0,010)	
	t-waarde	21,829	
AVI versie B	Parameter	0,218 ^b	0,893
	(Se)	(0,010)	
	t-waarde	21,829	
Lees en Begrijp 1 (Cito)	Parameter	5,965	0,617
	(Se)	(0,360)	
	t-waarde	16,570	
Lees en Begrijp 2 (Cito)	Parameter	16,633	0,694
	(Se)	(1,194)	
	t-waarde	13,925	
Begrijppentest Woorden (Aarmoutse)	Parameter	14,127	0,610
	(Se)	(0,872)	
	t-waarde	16,550	
Woordenschattest 3 (Aarmoutse)	Parameter	6,531	0,713
	(Se)	(0,508)	
	t-waarde	12,865	
COHORT 3 NAJAAR 1992			
Emt versie A	Parameter	6,567 ^c	0,965
	(Se)	(0,302)	
	t-waarde	21,772	
Emt versie B	Parameter	6,567 ^c	0,965
	(Se)	(0,302)	
	t-waarde	21,772	
AVI versie A	Parameter	-	1,000
	(Se)	-	
	t-waarde	-	
Lees en Begrijp 2 (Cito)	Parameter	13,972	0,700
	(Se)	(0,835)	
	t-waarde	16,724	
Begrijpend Leestest 4 (Aarmoutse)	Parameter	12,020	0,785
	(Se)	(0,907)	
	t-waarde	13,245	
Woordenschattest 3 (Aarmoutse)	Parameter	6,222	0,656
	(Se)	(0,357)	
	t-waarde	17,435	
Woordenschattest 4 (Aarmoutse)	Parameter	12,522	0,764
	(Se)	(0,931)	
	t-waarde	13,453	

		Theta-epsilon	R ²
COHORT 3 VOORJAAR 1993			
Emt versie A	Parameter	7,713 ^d	0,961
	(Se)	(0,355)	
	t-waarde	21,714	
Emt versie B	Parameter	7,713 ^d	0,961
	(Se)	(0,355)	
	t-waarde	21,714	
AVI versie A	Parameter	0,641 ^e	0,856
	(Se)	(0,030)	
	t-waarde	21,714	
AVI versie B	Parameter	0,641 ^e	0,856
	(Se)	(0,030)	
	t-waarde	21,714	
Begrijpend Lezen M3 (Cito)	Parameter	11,136	0,614
	(Se)	(0,662)	
	t-waarde	16,812	
Begrijpend Leestest 4 (Aarmoutse)	Parameter	10,815	0,722
	(Se)	(0,832)	
	t-waarde	13,003	
Woordenschattest 4 (Aarmoutse)	Parameter	6,690	0,814
	(Se)	(0,659)	
	t-waarde	10,151	
Woordenschattest 5 (Aarmoutse)	Parameter	14,357	0,689
	(Se)	(0,915)	
	t-waarde	15,683	
COHORT 4 VOORJAAR 1992			
Emt versie A	Parameter	9,670 ^f	0,955
	(Se)	(0,447)	
	t-waarde	21,633	
Emt versie B	Parameter	9,670 ^f	0,955
	(Se)	(0,447)	
	t-waarde	21,633	
AVI versie A	Parameter	0,523 ^g	0,905
	(Se)	(0,024)	
	t-waarde	21,633	
AVI versie B	Parameter	0,523 ^g	0,905
	(Se)	(0,024)	
	t-waarde	21,633	
Begrijpend Leestest 4 (Aarmoutse)	Parameter	12,914	0,654
	(Se)	(0,859)	
	t-waarde	15,036	
Begrijpend Lezen M3 (Cito)	Parameter	8,959	0,651
	(Se)	(0,592)	
	t-waarde	15,124	
Woordenschattest 4 (Aarmoutse)	Parameter	8,161	0,768
	(Se)	(0,703)	
	t-waarde	11,609	
Woordenschattest 5 (Aarmoutse)	Parameter	12,615	0,736
	(Se)	(0,971)	
	t-waarde	12,995	

		Theta-epsilon	R ²
COHORT 4 NAJAAR 1992			
Emt versie A	Parameter	12,493 ^h	0,943
	(Se)	(0,608)	
	t-waarde	20,555	
Emt versie B	Parameter	12,493 ^h	0,943
	(Se)	(0,608)	
	t-waarde	20,555	
AVI versie A	Parameter	-	1,000
	(Se)		
	t-waarde		
Begrijpend Lezen M3 (Cito)	Parameter	10,544	0,635
	(Se)	(0,665)	
	t-waarde	15,860	
Begrijpend Leestest 5 (Aarmoutse)	Parameter	9,470	0,780
	(Se)	(0,901)	
	t-waarde	10,510	
Woordenschattest 5 (Aarmoutse)	Parameter	13,559	0,715
	(Se)	(0,952)	
	t-waarde	14,243	
Woordenschattest 3 (Stijnen)	Parameter	14,508	0,805
	(Se)	(1,398)	
	t-waarde	10,378	
COHORT 4 VOORJAAR 1993			
Emt versie A	Parameter	11,018 ⁱ	0,944
	(Se)	(0,537)	
	t-waarde	20,518	
Emt versie B	Parameter	11,018 ⁱ	0,944
	(Se)	(0,537)	
	t-waarde	20,518	
AVI versie A	Parameter	0,696 ⁱ	0,851
	(Se)	(0,034)	
	t-waarde	20,518	
AVI versie B	Parameter	0,696 ⁱ	0,851
	(Se)	(0,034)	
	t-waarde	20,518	
Begrijpend Leestest 5 (Aarmoutse)	Parameter	8,804	0,727
	(Se)	(0,712)	
	t-waarde	12,361	
Begrijpend Lezen M4 (Cito)	Parameter	7,499	0,723
	(Se)	(0,599)	
	t-waarde	12,527	
Woordenschattest 3 (Stijnen)	Parameter	13,494	0,765
	(Se)	(1,451)	
	t-waarde	9,299	
Woordenschattest 6 (Aarmoutse)	Parameter	11,811	0,551
	(Se)	(0,717)	
	t-waarde	16,483	

		Theta-epsilon	R ²
COHORT 5 VOORJAAR 1992			
Emt versie A	Parameter	11,488 ^a	0,936
	(Se)	(0,557)	
	t-waarde	20,628	
Emt versie B	Parameter	11,488 ^a	0,936
	(Se)	(0,557)	
	t-waarde	20,628	
AVI versie A	Parameter	0,446 ^d	0,880
	(Se)	(0,022)	
	t-waarde	20,628	
AVI versie B	Parameter	0,446 ^d	0,880
	(Se)	(0,022)	
	t-waarde	20,628	
Begrijpend Leestest 5 (Aarmoutse)	Parameter	7,417	0,723
	(Se)	(0,579)	
	t-waarde	12,810	
Begrijpend Lezen M4 (Cito)	Parameter	5,763	0,722
	(Se)	(0,449)	
	t-waarde	12,830	
Woordenschattest 3 (Stijnen)	Parameter	33,905	0,488
	(Se)	(2,084)	
	t-waarde	16,273	
Woordenschattest 6 (Aarmoutse)	Parameter	12,689	0,532
	(Se)	(0,840)	
	t-waarde	15,111	
COHORT 5 NAJAAR 1992			
Emt versie A	Parameter	13,385 ^m	0,928
	(Se)	(0,667)	
	t-waarde	20,062	
Emt versie B	Parameter	13,385 ^m	0,928
	(Se)	(0,667)	
	t-waarde	20,062	
AVI versie A	Parameter	-	1,000
	(Se)		
	t-waarde		
Begrijpend Lezen M4 (Cito)	Parameter	7,160	0,665
	(Se)	(0,521)	
	t-waarde	13,746	
Begrijpend Leestest 6 (Aarmoutse)	Parameter	8,611	0,699
	(Se)	(0,686)	
	t-waarde	12,551	
Woordenschattest 6 (Aarmoutse)	Parameter	10,963	0,653
	(Se)	(0,748)	
	t-waarde	14,654	
Woordenschattest 4 (Stijnen)	Parameter	15,148	0,817
	(Se)	(1,840)	
	t-waarde	8,233	

		Theta-epsilon	R ²
COHORT 5 VOORJAAR 1993			
Emt versie A	Parameter	11,864 ⁿ	0,935
	(Se)	(0,593)	
	t-waarde	20,012	
Emt versie B	Parameter	11,864 ⁿ	0,935
	(Se)	(0,593)	
	t-waarde	20,012	
AVI versie A	Parameter	0,449 ^o	0,857
	(Se)	(0,022)	
	t-waarde	20,012	
AVI versie B	Parameter	0,449 ^o	0,857
	(Se)	(0,022)	
	t-waarde	20,012	
Begrijpend Leestest 6 (Aarmoutse)	Parameter	7,494	0,723
	(Se)	(0,592)	
	t-waarde	12,652	
Begrijpend Lezen E4 (Cito)	Parameter	5,513	0,697
	(Se)	(0,405)	
	t-waarde	13,619	
Woordenschattest 4 (Stijnen)	Parameter	10,282	0,841
	(Se)	(1,011)	
	t-waarde	10,167	
Woordenschattest 5 (Stijnen)	Parameter	10,996	0,848
	(Se)	(1,125)	
	t-waarde	9,778	

N.B.: De letters a t/m o geven aan welke parameters aan elkaar gelijkgesteld zijn

BIJLAGE 17 ONGESTANDAARDISEERDE PARAMETERSCHATTINGEN IN DE UITEINDELIJKE LONGITUDINALE MODELLEN VOOR DE RELATIE VAN AVI MET BEGRIJPEND LEZEN

Deze bijlage bevat de ongestandaardiseerde parameterschattingen voor de uiteindelijke longitudinale modellen voor de relatie tussen AVI en Begrijpend Lezen per cohort. De tabellen B17.1 tot en met B17.4 betreffen de parameterschattingen van model C2a van cohort 3. De tabellen B17.5 tot en met B17.8 betreffen de parameterschattingen van model C3 van cohort 4. Tot slot bevatten de tabellen B17.9 tot en met B17.12 de parameterschattingen van model D1 van cohort 5. Voor een beschrijving van de modellen verwijzen we naar tabel 4.13 in paragraaf 4.4.2.

Tabel B17.1: Factorladingen van de geobserveerde variabelen op de latente factoren voor cohort 3 in model C2a (Lambda-matrix)

		voorjaar groep 3		najaar groep 4		voorjaar groep 4	
		AVI ₁	BL ₁	AVI ₂	BL ₂	AVI ₃	BL ₃
COHORT 3 VOORJAAR 1992							
AVI versie A	Parameter	1,000					
	(Se)						
	t-waarde						
AVI versie B	Parameter	1,000					
	(Se)						
	t-waarde						
Lees en Begrijp 1 (Cito)	Parameter	1,000					
	(Se)						
	t-waarde						
Lees en Begrijp 2 (Cito)	Parameter	2,056 (0,087) 23,598					
	(Se)						
	t-waarde						
NAJAAR 1992							
AVI versie A	Parameter	1,000					
	(Se)						
	t-waarde						
Lees en Begrijp 2 (Cito)	Parameter	1,000					
	(Se)						
	t-waarde						
Begrijpend Leestest 4 (Aarmoutse)	Parameter	1,136 (0,037) 30,773					
	(Se)						
	t-waarde						

			voorjaar groep 3		najaar groep 4		voorjaar groep 4	
			AVI ₁	BL ₁	AVI ₂	BL ₂	AVI ₃	BL ₃
VOORJAAR 1993								
AVI versie A	Parameter						1,000	
	(Se)							
	t-waarde							
AVI versie B	Parameter						1,000	
	(Se)							
	t-waarde							
Begrijpend Lezen M3 (Cito)	Parameter							1,000
	(Se)							
	t-waarde							
Begrijpend Leestest 4 (Aarmoutse)	Parameter							1,211
	(Se)							(0,050)
	t-waarde							24,092

Tabel B17.2: Structurele relaties tussen de latente factoren voor cohort 3 in model C2a (Bèta-matrix)

onafhankelijke variabelen ->			voorjaar groep 3		najaar groep 4		voorjaar groep 4	
afhankelijke variabelen			AVI ₁	BL ₁	AVI ₂	BL ₂	AVI ₃	BL ₃
NAJAAR GROEP 4								
AVI ₂	Parameter		0,917	0,171				
	(Se)		(0,042)	(0,020)				
	t-waarde		22,088	8,409				
BL ₂	Parameter		-1,062 ^a	1,619	1,062 ^a			
	(Se)		(0,178)	(0,084)	(0,178)			
	t-waarde		-5,970	19,326	5,970			
VOORJAAR GROEP 4								
AVI ₃	Parameter				0,980			
	(Se)				(0,023)			
	t-waarde				41,993			
BL ₃	Parameter				-0,867 ^b	0,670	0,867 ^b	
	(Se)				(0,239)	(0,028)	(0,239)	
	t-waarde				-3,635	23,547	3,635	

N.B. De letters a en b geven aan welke parameters de restrictie is opgelegd dat ze een aan elkaar tegengestelde waarde hebben

Tabel B17.3: Varianties en covarianties van de residuen van de latente variabelen en de verklaarde varianties van de latente variabelen voor cohort 3 in model C2a (Psi-matrix)

		voorjaar groep 3		najaar groep 4		voorjaar groep 4	
		AVI ₁	BL ₁	AVI ₂	BL ₂	AVI ₃	BL ₃
COHORT 3 VOORJAAR GROEP 3							
AVI ₁	Parameter	1,811					
	(Se)	(0,093)					
	t-waarde	19,517					
BL ₁	Parameter	2,427	9,110				
	(Se)	(0,189)	(0,737)				
	t-waarde	12,834	12,360				
NAJAAR GROEP 4							
AVI ₂	Parameter			0,875			
	(Se)			(0,070)			
	t-waarde			12,542			
BL ₂	Parameter				4,321		
	(Se)				(0,773)		
	t-waarde				5,587		
VOORJAAR GROEP 4							
AVI ₃	Parameter					0,472	
	(Se)					(0,061)	
	t-waarde					7,762	
BL ₃	Parameter						2,978
	(Se)						(0,555)
	t-waarde						5,366
Verklaarde variantie (R ²)		-	-	0,745	0,873	0,875	0,839

Tabel B17.4: Error-varianties en -covarianties (Theta-epsilonmatrix) en verklaarde varianties (R^2) van de geobserveerde variabelen voor cohort 3 in model C2a

		Theta-epsilon	R ²
COHORT 3 VOORJAAR 1992			
AVI versie A	Parameter	0,217 ^a	0,893
	(Se)	(0,010)	
	t-waarde	20,721	
AVI versie B	Parameter	0,217 ^a	0,893
	(Se)	(0,010)	
	t-waarde	20,721	
Lees en Begrijp 1 (Cito)	Parameter	6,717	0,576
	(Se)	(0,412)	
	t-waarde	16,306	
Lees en Begrijp 2 (Cito)	Parameter	16,693	0,698
	(Se)	(1,292)	
	t-waarde	12,916	
NAJAAR 1992			
AVI versie A	Parameter	0,446	0,885
	(Se)	(0,051)	
	t-waarde	8,741	
Lees en Begrijp 2 (Cito)	Parameter	13,442	0,717
	(Se)	(0,933)	
	t-waarde	14,411	
Begrijpend Leestest 4 (Aarnoutse)	Parameter	12,406	0,780
	(Se)	(1,030)	
	t-waarde	12,050	
VOORJAAR 1993			
AVI versie A	Parameter	0,634 ^b	0,856
	(Se)	(0,031)	
	t-waarde	20,713	
AVI versie B	Parameter	0,634 ^b	0,856
	(Se)	(0,031)	
	t-waarde	20,713	
Begrijpend Lezen M3 (Cito)	Parameter	10,748	0,632
	(Se)	(0,744)	
	t-waarde	14,447	
Begrijpend Leestest 4 (Aarnoutse)	Parameter	12,462	0,685
	(Se)	(0,979)	
	t-waarde	12,733	
Error-covariantie tussen Lees en Begrijp 2 in het voorjaar 1992 en in het najaar 1992	Parameter	2,252	
	(Se)	(0,808)	
	t-waarde	2,787	
Error-covariantie tussen Begrijpend Leestest 4 in het najaar 1992 en in het voorjaar 1993	Parameter	1,498	
	(Se)	(0,733)	
	t-waarde	2,043	

N B De letters a en b geven aan welke parameters aan elkaar gelijkgesteld zijn

Tabel B17.5: Factorladingen van de geobserveerde variabelen op de latente factoren voor cohort 4 in model C3 (Lambda-matrix)

		voorjaar groep 4		najaar groep 5		voorjaar groep 5	
		AVI ₁	BL ₁	AVI ₂	BL ₂	AVI ₃	BL ₃
COHORT 4 VOORJAAR 1992							
AVI versie A	Parameter	1,000					
	(Se)						
	t-waarde						
AVI versie B	Parameter	1,000					
	(Se)						
	t-waarde						
Begrijpend Leestest 4 (Aarmoutse)	Parameter		1,000				
	(Se)						
	t-waarde						
Begrijpend Lezen M3 (Cito)	Parameter		0,826				
	(Se)		(0,034)				
	t-waarde		24,558				
NAJAAR 1992							
AVI versie A	Parameter			1,000			
	(Se)						
	t-waarde						
Begrijpend Lezen M3 (Cito)	Parameter				1,000		
	(Se)						
	t-waarde						
Begrijpend Leestest 5 (Aarmoutse)	Parameter				1,269		
	(Se)				(0,049)		
	t-waarde				26,080		
VOORJAAR 1993							
AVI versie A	Parameter					1,000	
	(Se)						
	t-waarde						
AVI versie B	Parameter					1,000	
	(Se)						
	t-waarde						
Begrijpend Leestest 5 (Aarmoutse)	Parameter						1,000
	(Se)						
	t-waarde						
Begrijpend Lezen M4 (Cito)	Parameter						0,898
	(Se)						(0,034)
	t-waarde						26,714

Tabel B17.6: Structurele relaties tussen de latente factoren voor cohort 4 in model C3 (Bèta-matrix)

onafhankelijke variabelen ->		voorjaar groep 4		najaar groep 5		voorjaar groep 5	
afhankelijke variabelen		AVI ₁	BL ₁	AVI ₂	BL ₂	AVI ₁	BL ₁
NAJAAR GROEP 5							
AVI ₂	Parameter	0,963					
	(Se)	(0,019)					
	t-waarde	50,678					
BL ₂	Parameter	-0,846 ^a	0,866	0,846 ^a			
	(Se)	(0,254)	(0,039)	(0,254)			
	t-waarde	-3,325	22,436	3,325			
VOORJAAR GROEP 5							
AVI ₃	Parameter			0,895			
	(Se)			(0,018)			
	t-waarde			50,756			
BL ₃	Parameter				1,005		
	(Se)				(0,043)		
	t-waarde				23,188		

N.B De letter a geeft aan welke parameters de restrictie is opgelegd dat ze een aan elkaar tegengestelde waarde hebben

Tabel B17.7: Varianties en covarianties van de residuen van de latente variabelen en de verklaarde varianties van de latente variabelen voor cohort 4 in model C3 (Psi-matrix)

		voorjaar groep 4		najaar groep 5		voorjaar groep 5	
		AVI ₁	BL ₁	AVI ₂	BL ₂	AVI ₁	BL ₁
COHORT 4 VOORJAAR GROEP 4							
AVI ₁	Parameter	4,651					
	(Se)	(0,247)					
	t-waarde	18,857					
BL ₁	Parameter	6,982	23,976				
	(Se)	(0,508)	(1,794)				
	t-waarde	13,731	13,365				
NAJAAR GROEP 5							
AVI ₂	Parameter			0,370			
	(Se)			(0,051)			
	t-waarde			7,260			
BL ₂	Parameter				0,987		
	(Se)				(0,462)		
	t-waarde				2,135		
VOORJAAR GROEP 5							
AVI ₃	Parameter					0,165	
	(Se)					(0,044)	
	t-waarde					3,724	
BL ₃	Parameter						3,925
	(Se)						(0,625)
	t-waarde						6,279
Verklaarde variantie (R²)		-	-	0,921	0,948	0,958	0,829

Tabel B17.8: Error-varianties en -covarianties (Theta-epsilonmatrix) en verklaarde varianties (R^2) van de geobserveerde variabelen voor cohort 4 in model C3

		Theta-epsilon	R ²
COHORT 4 VOORJAAR 1992			
AVI versie A	Parameter	0,524 ^a	0,899
	(Se)	(0,026)	
	t-waarde	20,092	
AVI versie B	Parameter	0,524 ^a	0,899
	(Se)	(0,026)	
	t-waarde	20,092	
Begrijpend Leestest 4 (Aarmoutse)	Parameter	12,003	0,666
	(Se)	(0,828)	
	t-waarde	14,488	
Begrijpend Lezen M3 (Cito)	Parameter	9,262	0,639
	(Se)	(0,613)	
	t-waarde	15,114	
NAJAAR 1992			
AVI versie A	Parameter	0,643	0,879
	(Se)	(0,051)	
	t-waarde	12,594	
Begrijpend Lezen M3 (Cito)	Parameter	9,967	0,654
	(Se)	(0,651)	
	t-waarde	15,317	
Begrijpend Leestest 5 (Aarmoutse)	Parameter	12,314	0,712
	(Se)	(0,894)	
	t-waarde	13,779	
VOORJAAR 1993			
AVI versie A	Parameter	0,695 ^b	0,850
	(Se)	(0,035)	
	t-waarde	19,937	
AVI versie B	Parameter	0,695 ^b	0,850
	(Se)	(0,035)	
	t-waarde	19,937	
Begrijpend Leestest 5 (Aarmoutse)	Parameter	8,269	0,735
	(Se)	(0,704)	
	t-waarde	11,751	
Begrijpend Lezen M4 (Cito)	Parameter	7,717	0,706
	(Se)	(0,602)	
	t-waarde	12,810	
Error-covariantie tussen Begrijpen Lezen M3 in het voorjaar 1992 en in het najaar 1992	Parameter	1,218	
	(Se)	(0,478)	
	t-waarde	2,549	
Error-covariantie tussen Begrijpend Leestest 5 in het najaar 1992 en in het voorjaar 1993	Parameter	0,697	
	(Se)	(0,577)	
	t-waarde	1,209	

N B De letters a en b geven aan welke parameters aan elkaar gelijkgesteld zijn

Tabel B17.9: Factorloadingen van de geobserveerde variabelen op de latente factoren voor cohort 5 in model D1 (Lambda-matrix)

		voorjaar groep 5		najaar groep 6		voorjaar groep 6	
		AVI ₁	BL ₁	AVI ₂	BL ₂	AVI ₃	BL ₃
COHORT 5 VOORJAAR 1992							
AVI versie A	Parameter	1,000					
	(Se)						
	t-waarde						
	Parameter	1,000					
AVI versie B	(Se)						
	t-waarde						
	Parameter		1,000				
Begrijpend Leestest 5 (Aarmoutse)	(Se)						
	t-waarde						
	Parameter		0,849				
Begrijpend Lezen M4 (Cito)	(Se)			(0,030)			
	t-waarde		28,033				
	NAJAAR 1992						
AVI versie A	Parameter			1,000			
	(Se)						
	t-waarde						
	Parameter				1,000		
Begrijpend Lezen M4 (Cito)	(Se)						
	t-waarde						
	Parameter				1,106		
Begrijpend Leestest 6 (Aarmoutse)	(Se)					(0,043)	
	t-waarde				25,914		
	VOORJAAR 1993						
AVI versie A	Parameter					1,000	
	(Se)						
	t-waarde						
	Parameter					1,000	
AVI versie B	(Se)						
	t-waarde						
	Parameter						1,000
Begrijpend Leestest 6 (Aarmoutse)	(Se)						
	t-waarde						
	Parameter						0,784
Begrijpend Lezen E4 (Cito)	(Se)						
	t-waarde						26,481

Tabel B17.10: Structurele relaties tussen de latente factoren voor cohort 5 in model D1 (Bèta-matrix)

onafhankelijke variabelen ->		voorjaar groep 5		najaar groep 6		voorjaar groep 6	
afhankelijke variabelen		AVI ₁	BL ₁	AVI ₂	BL ₂	AVI ₃	BL ₃
NAJAAR GROEP 6 AVI ₂	Parameter	0,917					
	(Se)	(0,021)					
	t-waarde	43,006					
BL ₂	Parameter		0,865				
	(Se)		(0,033)				
	t-waarde		26,231				
VOORJAAR GROEP 6							
AVI ₃	Parameter			0,868			
	(Se)			(0,021)			
	t-waarde			41,878			
BL ₃	Parameter				1,069		
	(Se)				(0,042)		
	t-waarde				25,463		

Tabel B17.11: Varianties en covarianties van de residuen van de latente variabelen en de verklaarde varianties van de latente variabelen voor cohort 5 in model D1 (Psi-matrix)

		voorjaar groep 5		najaar groep 6		voorjaar groep 6	
		AVI ₁	BL ₁	AVI ₂	BL ₂	AVI ₃	BL ₃
COHORT 5 VOORJAAR GROEP 5							
AVI ₁	Parameter	3,276					
	(Se)	(0,182)					
	t-waarde	17,991					
BL ₁	Parameter	4,689	20,519				
	(Se)	(0,381)	(1,444)				
	t-waarde	12,316	14,214				
NAJAAR GROEP 6							
AVI ₂	Parameter			0,337			
	(Se)			(0,045)			
	t-waarde			7,494			
BL ₂	Parameter				0,184		
	(Se)				(0,327)		
	t-waarde				0,561		
VOORJAAR GROEP 6							
AVI ₃	Parameter					0,238	
	(Se)					(0,038)	
	t-waarde					6,321	
BL ₃	Parameter						2,522
	(Se)						(0,490)
	t-waarde						5,147
Verklaarde variantie (R ²)		-	-	0,891	0,988	0,907	0,876

Tabel B17.12: Error-varianties en -covarianties (Theta-epsilon matrix) en verklaarde varianties (R^2) van de geobserveerde variabelen voor cohort 5 in model D1

		Theta-epsilon	R ²
COHORT 5 VOORJAAR 1992			
AVI versie A	Parameter	0,431 ^a	0,884
	(Se)	(0,022)	
	t-waarde	19,329	
AVI versie B	Parameter	0,431 ^a	0,884
	(Se)	(0,022)	
	t-waarde	19,329	
Begrijpend Leestest 5 (Aamoutse)	Parameter	7,013	0,745
	(Se)	(0,544)	
	t-waarde	12,886	
Begrijpend Lezen M4 (Cito)	Parameter	6,307	0,701
	(Se)	(0,445)	
	t-waarde	14,159	
NAJAAR 1992			
AVI versie A	Parameter	0,484	0,865
	(Se)	(0,044)	
	t-waarde	11,053	
Begrijpend Lezen M4 (Cito)	Parameter	6,188	0,715
	(Se)	(0,443)	
	t-waarde	13,977	
Begrijpend Leestest 6 (Aamoutse)	Parameter	9,988	0,656
	(Se)	(0,650)	
	t-waarde	15,373	
VOORJAAR 1993			
AVI versie A	Parameter	0,436 ^b	0,855
	(Se)	(0,023)	
	t-waarde	19,209	
AVI versie B	Parameter	0,436 ^b	0,855
	(Se)	(0,023)	
	t-waarde	19,209	
Begrijpend Leestest 6 (Aamoutse)	Parameter	7,121	0,740
	(Se)	(0,596)	
	t-waarde	11,952	
Begrijpend Lezen E4 (Cito)	Parameter	5,782	0,683
	(Se)	(0,419)	
	t-waarde	13,797	
Error-covariantie tussen Begrijpen Lezen M4 in het voorjaar 1992 en in het najaar 1992	Parameter	0,616	
	(Se)	(0,353)	
	t-waarde	1,743	
Error-covariantie tussen Begrijpend Leestest 6 in het najaar 1992 en in het voorjaar 1993	Parameter	2,297	
	(Se)	(0,462)	
	t-waarde	4,970	

N.B.. De letters a en b geven aan welke parameters aan elkaar gelijkgesteld zijn

BIJLAGE 18 ONGESTANDAARDISEERDE PARAMETERSCHATTINGEN IN DE UITEINDELIJKE LONGITUDINALE MODELLEN VOOR DE RELATIE VAN AVI MET WOORDENSCHAT

Deze bijlage bevat de ongestandaardiseerde parameterschattingen voor de uiteindelijke longitudinale modellen voor de relatie tussen AVI en Woordenschat per cohort. De tabellen B18.1 tot en met B18.4 betreffen de parameterschattingen van model C2a van cohort 3. De tabellen B18.5 tot en met B18.8 betreffen de parameterschattingen van model C3 van cohort 4. Tot slot bevatten de tabellen B18.9 tot en met B18.12 de parameterschattingen van model D1 van cohort 5. Voor een beschrijving van de modellen verwijzen we naar tabel 4.14 in paragraaf 4.4.3.

Tabel B18.1: Factorladingen van de geobserveerde variabelen op de latente factoren voor cohort 3 in model C2a (Lambda-matrix)

			voorjaar groep 3		najaar groep 4		voorjaar groep 4	
			AVI ₁	WS ₁	AVI ₂	WS ₂	AVI ₃	WS ₃
COHORT 3 VOORJAAR 1992								
AVI versie A	Parameter		1,000					
	(Se)							
	t-waarde							
AVI versie B	Parameter		1,000					
	(Se)							
	t-waarde							
Begrippentest Woorden (Aarmoutse)	Parameter			1,000				
	(Se)							
	t-waarde							
Woordenschattest 3 (Aarmoutse)	Parameter			0,842				
	(Se)			(0,037)				
	t-waarde			22,635				
NAJAAR 1992								
AVI versie A	Parameter				1,000			
	(Se)							
	t-waarde							
Woordenschattest 3 (Aarmoutse)	Parameter					1,000		
	(Se)							
	t-waarde							
Woordenschattest 4 (Aarmoutse)	Parameter					2,005		
	(Se)					(0,070)		
	t-waarde					28,661		

			voorjaar groep 3		najaar groep 4		voorjaar groep 4	
			AVI ₁	WS ₁	AVI ₂	WS ₂	AVI ₃	WS ₃
VOORJAAR 1993								
AVI versie A	Parameter						1,000	
	(Se)							
	t-waarde							
AVI versie B	Parameter						1,000	
	(Se)							
	t-waarde							
Woordenschattest 4 (Aamoutse)	Parameter							1,000
	(Se)							
	t-waarde							
Woordenschattest 5 (Aamoutse)	Parameter							1,028
	(Se)							(0,034)
	t-waarde							30,501

Tabel B18.2: Structurele relaties tussen de latente factoren voor cohort 3 in model C2a (Bèta-matrix)

onafhankelijke variabelen ->			voorjaar groep 3		najaar groep 4		voorjaar groep 4	
afhankelijke variabelen			AVI ₁	WS ₁	AVI ₂	WS ₂	AVI ₃	WS ₃
NAJAAR GROEP 4								
AVI ₂	Parameter		1,014	0,068				
	(Se)		(0,044)	(0,014)				
	t-waarde		23,191	5,011				
WS ₂	Parameter		-0,821 ^a	0,559	0,821 ^a			
	(Se)		(0,102)	(0,031)	(0,102)			
	t-waarde		-8,028	17,806	8,028			
VOORJAAR GROEP 4								
AVI ₃	Parameter				0,976			
	(Se)				(0,024)			
	t-waarde				41,176			
WS ₃	Parameter				-0,442 ^b	1,550	0,442 ^b	
	(Se)				(0,252)	(0,060)	(0,252)	
	t-waarde				-1,754	25,782	1,754	

N.B. De letters a en b geven aan welke parameters de restrictie is opgelegd dat ze een aan elkaar tegengestelde waarde hebben

Tabel B18.3: Varianties en covarianties van de residuen van de latente variabelen en de verklaarde varianties van de latente variabelen voor cohort 3 in model C2a (Psi-matrix)

		voorjaar groep 3		najaar groep 4		voorjaar groep 4	
		AVI ₁	WS ₁	AVI ₂	WS ₂	AVI ₃	WS ₃
COHORT 3 VOORJAAR GROEP 3							
AVI ₁	Parameter	1,803					
	(Se)	(0,092)					
	t-waarde	19,506					
WS ₁	Parameter	3,794	21,786				
	(Se)	(0,294)	(1,772)				
	t-waarde	12,901	12,295				
NAJAAR GROEP 4							
AVI ₂	Parameter			0,976			
	(Se)			(0,075)			
	t-waarde			13,102			
WS ₂	Parameter				2,521		
	(Se)				(0,295)		
	t-waarde				8,536		
VOORJAAR GROEP 4							
AVI ₃	Parameter					0,478	
	(Se)					(0,062)	
	t-waarde					7,696	
WS ₃	Parameter						2,745
	(Se)						(0,688)
	t-waarde						3,991
Verklaarde variantie (R ²)		-	-	0,717	0,780	0,873	0,909

Tabel B18.4: Error-varianties en -covarianties (Theta-epsilonmatrix) en verklaarde varianties (R^2) van de geobserveerde variabelen voor cohort 3 in model C2a

		Theta-epsilon	R ²
COHORT 3 VOORJAAR 1992			
AVI versie A	Parameter	0,218 ^a	0,892
	(Se)	(0,011)	
	t-waarde	20,731	
AVI versie B	Parameter	0,218 ^a	0,892
	(Se)	(0,011)	
	t-waarde	20,731	
Begnppentest Woorden (Aamoutse)	Parameter	15,678	0,582
	(Se)	(1,007)	
	t-waarde	15,570	
Woordenschattest 3 (Aamoutse)	Parameter	7,480	0,674
	(Se)	(0,578)	
	t-waarde	12,947	
NAJAAR 1992			
AVI versie A	Parameter	0,444	0,886
	(Se)	(0,053)	
	t-waarde	8,356	
Woordenschattest 3 (Aamoutse)	Parameter	7,241	0,613
	(Se)	(0,415)	
	t-waarde	17,455	
Woordenschattest 4 (Aamoutse)	Parameter	8,574	0,843
	(Se)	(0,950)	
	t-waarde	9,026	
VOORJAAR 1993			
AVI versie A	Parameter	0,634 ^b	0,856
	(Se)	(0,031)	
	t-waarde	20,712	
AVI versie B	Parameter	0,634 ^b	0,856
	(Se)	(0,031)	
	t-waarde	20,712	
Woordenschattest 4 (Aamoutse)	Parameter	6,328	0,827
	(Se)	(0,693)	
	t-waarde	9,137	
Woordenschattest 5 (Aamoutse)	Parameter	14,452	0,688
	(Se)	(0,946)	
	t-waarde	15,284	
Error-covariantie tussen Woordenschattest 3 in het voorjaar 1992 en in het najaar 1992	Parameter	1,441	
	(Se)	(0,341)	
	t-waarde	4,226	
Error-covariantie tussen Woordenschattest 4 in het najaar 1992 en in het voorjaar 1993	Parameter	-1,866	
	(Se)	(0,650)	
	t-waarde	-2,870	

N B De letters a en b geven aan welke parameters aan elkaar gelijkgesteld zijn

Tabel B18.5: Factorladingen van de geobserveerde variabelen op de latente factoren voor cohort 4 in model C3 (Lambda-matrix)

		voorjaar groep 4		najaar groep 5		voorjaar groep 5	
		AVI ₁	WS ₁	AVI ₂	WS ₂	AVI ₃	WS ₃
COHORT 4 VOORJAAR 1992							
AVI versie A	Parameter	1,000					
	(Se)						
	t-waarde						
AVI versie B	Parameter	1,000					
	(Se)						
	t-waarde						
Woordenschattest 4 (Aarmoutse)	Parameter		1,000				
	(Se)						
	t-waarde						
Woordenschattest 5 (Aarmoutse)	Parameter		1,233				
	(Se)		(0,040)				
	t-waarde		31,051				
NAJAAR 1992							
AVI versie A	Parameter			1,000			
	(Se)						
	t-waarde						
Woordenschattest 5 (Aarmoutse)	Parameter				1,000		
	(Se)						
	t-waarde						
Woordenschattest 3 (Stijnen)	Parameter				1,197		
	(Se)				(0,036)		
	t-waarde				33,132		
VOORJAAR 1993							
AVI versie A	Parameter					1,000	
	(Se)						
	t-waarde						
AVI versie B	Parameter					1,000	
	(Se)						
	t-waarde						
Woordenschattest 3 (Stijnen)	Parameter						1,000
	(Se)						
	t-waarde						
Woordenschattest 6 (Aarmoutse)	Parameter						0,683
	(Se)						(0,027)
	t-waarde						25,298

Tabel B18.6: Structurele relaties tussen de latente factoren voor cohort 4 in model C3 (Bèta-matrix)

onafhankelijke variabelen ->		voorjaar groep 4		najaar groep 5		voorjaar groep 5	
afhankelijke variabelen		AVI ₁	WS ₁	AVI ₂	WS ₂	AVI ₃	WS ₃
NAJAAR GROEP 5	Parameter	0,970					
	(Se)	(0,019)					
	t-waarde	50,812					
WS ₂	Parameter	-0,808 ^a	1,234	0,808 ^a			
	(Se)	(0,277)	(0,042)	(0,277)			
	t-waarde	-2,915	29,664	2,915			
VOORJAAR GROEP 5							
AVI ₃	Parameter			0,890			
	(Se)			(0,017)			
	t-waarde			50,871			
WS ₃	Parameter				0,959		
	(Se)				(0,035)		
	t-waarde				27,659		

N B De letter a geeft aan welke parameters de restrictie is opgelegd dat ze een aan elkaar tegengestelde waarde hebben

Tabel B18.7: Varianties en covarianties van de residuen van de latente variabelen en de verklaarde varianties van de latente variabelen voor cohort 4 in model C3 (Psi-matrix)

		voorjaar groep 4		najaar groep 5		voorjaar groep 5	
		AVI ₁	WS ₁	AVI ₂	WS ₂	AVI ₃	WS ₃
COHORT 4 VOORJAAR GROEP 4							
AVI ₁	Parameter	4,633					
	(Se)	(0,246)					
	t-waarde	18,857					
WS ₁	Parameter	6,365	24,348				
	(Se)	(0,481)	(1,673)				
	t-waarde	13,225)	14,558				
NAJAAR GROEP 5							
AVI ₂	Parameter			0,382			
	(Se)			(0,052)			
	t-waarde			7,361			
WS ₂	Parameter				1,286		
	(Se)				(0,599)		
	t-waarde				2,146		
VOORJAAR GROEP 5							
AVI ₃	Parameter					0,169	
	(Se)					(0,045)	
	t-waarde					3,791	
WS ₃	Parameter						1,662
	(Se)						(0,869)
	t-waarde						1,913
Verklaarde variantie (R ²)		-	-	0,919	0,966	0,957	0,955

Tabel B18.8: Error-varianties en -covarianties (Theta-epsilon-matrix) en verklaarde varianties (R^2) van de geobserveerde variabelen voor cohort 4 in model C3

		Theta-epsilon	R ²
COHORT 4 VOORJAAR 1992			
AVI versie A	Parameter	0,523 ^a	0,899
	(Se)	(0,026)	
	t-waarde	20,070	
AVI versie B	Parameter	0,523 ^a	0,899
	(Se)	(0,026)	
	t-waarde	20,070	
Woordenschattest 4 (Aarmoutse)	Parameter	9,185	0,726
	(Se)	(0,627)	
	t-waarde	14,653	
Woordenschattest 5 (Aarmoutse)	Parameter	11,485	0,763
	(Se)	(0,854)	
	t-waarde	13,447	
NAJAAR 1992			
AVI versie A	Parameter	0,636	0,882
	(Se)	(0,051)	
	t-waarde	12,381	
Woordenschattest 5 (Aarmoutse)	Parameter	9,865	0,795
	(Se)	(0,757)	
	t-waarde	13,026	
Woordenschattest 3 (Stijnen)	Parameter	18,660	0,746
	(Se)	(1,277)	
	t-waarde	14,608	
VOORJAAR 1993			
AVI versie A	Parameter	0,695 ^b	0,850
	(Se)	(0,035)	
	t-waarde	19,937	
AVI versie B	Parameter	0,695 ^b	0,850
	(Se)	(0,035)	
	t-waarde	19,937	
Woordenschattest 3 (Stijnen)	Parameter	19,504	0,653
	(Se)	(1,313)	
	t-waarde	14,859	
Woordenschattest 6 (Aarmoutse)	Parameter	8,958	0,657
	(Se)	(0,614)	
	t-waarde	14,581	
Error-covariantie tussen Woordenschattest 5 van Aarmoutse in het voorjaar 1992 en in het najaar 1992	Parameter	3,636	
	(Se)	(0,660)	
	t-waarde	5,507	
Error-covariantie tussen Woordenschattest 3 van Stijnen in het najaar 1992 en in het voorjaar 1993	Parameter	3,972	
	(Se)	(0,974)	
	t-waarde	4,080	

N.B. De letters a en b geven aan welke parameters aan elkaar gelijkgesteld zijn

Tabel B18.9: Factorladingen van de geobserveerde variabelen op de latente factoren voor cohort 5 in model D1 (Lambda-matrix)

			voorjaar groep 5		najaar groep 6		voorjaar groep 6	
			AVI ₁	WS ₁	AVI ₂	WS ₂	AVI ₃	WS ₃
COHORT 5 VOORJAAR 1992								
AVI versie A	Parameter		1,000					
	(Se)							
	t-waarde							
AVI versie B	Parameter		1,000					
	(Se)							
	t-waarde							
Woordenschattest 3 (Stijnen)	Parameter			1,000				
	(Se)							
	t-waarde							
Woordenschattest 6 (Aarmoutse)	Parameter			0,700				
	(Se)			(0,035)				
	t-waarde			20,121				
NAJAAR 1992								
AVI versie A	Parameter				1,000			
	(Se)							
	t-waarde							
Woordenschattest 6 (Aarmoutse)	Parameter					1,000		
	(Se)							
	t-waarde							
Woordenschattest 4 (Stijnen)	Parameter					1,870		
	(Se)					(0,066)		
	t-waarde					28,378		
VOORJAAR 1993								
AVI versie A	Parameter						1,000	
	(Se)							
	t-waarde							
AVI versie B	Parameter						1,000	
	(Se)							
	t-waarde							
Woordenschattest 4 (Stijnen)	Parameter							1,000
	(Se)							
	t-waarde							
Woordenschattest 5 (Stijnen)	Parameter							1,008
	(Se)							(0,026)
	t-waarde							39,270

Tabel B18.10: Structurele relaties tussen de latente factoren voor cohort 5 in model D1 (Bèta-matrix)

onafhankelijke variabelen ->		voorjaar groep 5		najaar groep 6		voorjaar groep 6	
afhankelijke variabelen		AVI ₁	WS ₁	AVI ₂	WS ₂	AVI ₃	WS ₃
NAJAAR GROEP 6							
AVI ₂	Parameter	0,916					
	(Se)	(0,021)					
	t-waarde	42,883					
WS ₂	Parameter		0,777				
	(Se)		(0,038)				
	t-waarde		20,678				
VOORJAAR GROEP 6							
AVI ₃	Parameter			0,867			
	(Se)			(0,021)			
	t-waarde			41,882			
WS ₃	Parameter				1,639		
	(Se)				(0,060)		
	t-waarde				27,195		

Tabel B18.11: Varianties en covarianties van de residuen van de latente variabelen en de verklaarde varianties van de latente variabelen voor cohort 5 in model D1 (Psi-matrix)

		voorjaar groep 5		najaar groep 6		voorjaar groep 6	
		AVI ₁	WS ₁	AVI ₂	WS ₂	AVI ₃	WS ₃
COHORT 5 VOORJAAR GROEP 5							
AVI ₁	Parameter	3,279					
	(Se)	(0,182)					
	t-waarde	18,002					
WS ₁	Parameter	5,663	32,750				
	(Se)	(0,502)	(3,037)				
	t-waarde	11,277	10,784				
NAJAAR GROEP 6							
AVI ₂	Parameter			0,346			
	(Se)			(0,045)			
	t-waarde			7,623			
WS ₂	Parameter				0,001*		
	(Se)						
	t-waarde						
VOORJAAR GROEP 6							
AVI ₃	Parameter					0,240	
	(Se)					(0,038)	
	t-waarde					6,339	
WS ₃	Parameter						4,370
	(Se)						(0,984)
	t-waarde						4,441
Verklaarde variantie (R²)		-	-	0,888	1,000	0,907	0,924

NB De letter a geeft aan dat deze parameter is vastgezet op de waarde 0,001

Tabel B18.12: Error-varianties en -covarianties (Theta-epsilon matrix) en verklaarde varianties (R^2) van de geobserveerde variabelen voor cohort 5 in model D1

		Theta-epsilon	R ²
COHORT 5 VOORJAAR 1992			
AVI versie A	Parameter	0,427 ^a	0,885
	(Se)	(0,022)	
	t-waarde	19,316	
AVI versie B	Parameter	0,427 ^a	0,885
	(Se)	(0,022)	
	t-waarde	19,316	
Woordenschattest 3 (Stijnen)	Parameter	33,024	0,498
	(Se)	(1,847)	
	t-waarde	17,883	
Woordenschattest 6 (Aarmoutse)	Parameter	11,050	0,592
	(Se)	(0,650)	
	t-waarde	17,007	
NAJAAR 1992			
AVI versie A	Parameter	0,482	0,865
	(Se)	(0,044)	
	t-waarde	10,975	
Woordenschattest 6 (Aarmoutse)	Parameter	11,830	0,626
	(Se)	(0,710)	
	t-waarde	16,660	
Woordenschattest 4 (Stijnen)	Parameter	12,423	0,848
	(Se)	(1,275)	
	t-waarde	9,743	
VOORJAAR 1993			
AVI versie A	Parameter	0,436 ^b	0,855
	(Se)	(0,023)	
	t-waarde	19,209	
AVI versie B	Parameter	0,436 ^b	0,855
	(Se)	(0,023)	
	t-waarde	19,209	
Woordenschattest 4 (Stijnen)	Parameter	6,653	0,896
	(Se)	(0,900)	
	t-waarde	7,391	
Woordenschattest 5 (Stijnen)	Parameter	13,901	0,808
	(Se)	(1,105)	
	t-waarde	12,581	
Error-covariantie tussen Woordenschattest 6 van Aarmoutse in het voorjaar 1992 en in het najaar 1992	Parameter	5,265	
	(Se)	(0,548)	
	t-waarde	9,608	
Error-covariantie tussen Woordenschattest 4 van Stijnen in het najaar 1992 en in het voorjaar 1993	Parameter	-1,272	
	(Se)	(0,856)	
	t-waarde	-1,486	

N.B.: De letters a en b geven aan welke parameters aan elkaar gelijkgesteld zijn

BIJLAGE 19: CORRELATIEMATRICES, GEMIDDELDEN EN STANDAARDAFWIJKINGEN VOOR LISREL-ANALYSES

Tabel B19.1: Correlaties tussen geobserveerde variabelen en gemiddelden en standaardafwijkingen van de variabelen in cohort 3 periode 1 (voorjaar 1992; N=954)

	CLB1	CLB2	BEGW	AWS3	EMTA	EMTB	AVIA	AVIB
Lees en Begrijp 1 (Cito) CLB1	1,000	0,654	0,611	0,630	0,469	0,458	0,389	0,399
Lees en Begrijp 2 (Cito) CLB2	0,654	1,000	0,602	0,681	0,532	0,523	0,443	0,471
Begrijppentest Woorden(Aarmoutse) BEGW	0,611	0,602	1,000	0,660	0,455	0,438	0,410	0,423
Woordenschattest 3 (Aarmoutse) AWS3	0,630	0,681	0,660	1,000	0,465	0,472	0,424	0,445
Emt versie A EMTA	0,469	0,532	0,455	0,465	1,000	0,956	0,782	0,788
Emt versie B EMTB	0,458	0,523	0,438	0,472	0,956	1,000	0,791	0,801
AVI versie A AVIA	0,389	0,443	0,410	0,424	0,782	0,791	1,000	0,893
AVI versie B AVIB	0,399	0,471	0,423	0,445	0,788	0,801	0,893	1,000
Gemiddelde	22,8	21,5	28,9	19,4	26,3	24,8	1,2	1,2
Standaardafwijking	3,9	7,4	6,1	4,8	12,0	12,0	1,4	1,4

Tabel B19.2: Correlaties tussen geobserveerde variabelen en gemiddelden en standaardafwijkingen van de variabelen in cohort 3 periode 2 (najaar 1992; N=949)

	CLB2	ABL4	AWS3	AWS4	EMTA	EMTB	AVIA
Lees en Begrijp 2 (Cito) CLB2	1,000	0,741	0,672	0,711	0,607	0,591	0,572
Begrijpend Leestest 4 (Aarmoutse) ABL4	0,741	1,000	0,700	0,765	0,616	0,618	0,611
Woordenschattest 3 (Aarmoutse) AWS3	0,672	0,700	1,000	0,708	0,576	0,566	0,558
Woordenschattest 4 (Aarmoutse) AWS4	0,711	0,765	0,708	1,000	0,617	0,613	0,607
Emt versie A EMTA	0,607	0,616	0,576	0,617	1,000	0,965	0,846
Emt versie B EMTB	0,591	0,618	0,566	0,613	0,965	1,000	0,845
AVI versie A AVIA	0,572	0,611	0,558	0,607	0,846	0,845	1,000
Gemiddelde	22,4	21,3	21,4	21,6	31,6	30,5	3,0
Standaardafwijking	6,8	7,5	4,3	7,3	13,5	13,9	2,0

Tabel B19.3: Correlaties tussen geobserveerde variabelen en gemiddelden en standaardafwijkingen van de variabelen in cohort 3 periode 3 (voorjaar 1993; N=944)

	CBM3	ABL4	AWS4	AWS5	EMTA	EMTB	AVIA	AVIB
Begrijpend Lezen M3 (Cito) CBM3	1,000	0,666	0,631	0,620	0,501	0,498	0,479	0,496
Begrijpend Leestest 4(Aarmoutse) ABL4	0,666	1,000	0,724	0,645	0,481	0,481	0,479	0,487
Woordenschattest 4 (Aarmoutse) AWS4	0,631	0,724	1,000	0,749	0,508	0,513	0,501	0,509
Woordenschattest 5 (Aarmoutse) AWS5	0,620	0,645	0,749	1,000	0,431	0,433	0,443	0,477
Emt versie A EMTA	0,501	0,481	0,508	0,431	1,000	0,962	0,799	0,817
Emt versie B EMTB	0,498	0,481	0,513	0,433	0,962	1,000	0,804	0,817
AVI versie A AVIA	0,479	0,479	0,501	0,443	0,799	0,804	1,000	0,856
AVI versie B AVIB	0,496	0,487	0,509	0,477	0,817	0,817	0,856	1,000
Gemiddelde	16,5	26,8	27,1	21,0	43,0	42,0	4,5	4,5
Standaardafwijking	5,4	6,2	6,0	6,8	13,9	14,3	2,1	2,1

Tabel B19.4: Correlaties tussen geobserveerde variabelen en gemiddelden en standaardafwijkingen van de variabelen in cohort 4 periode 1 (voorjaar 1992; N=937)

	ABL4	CBM3	AWS4	AWS5	EMTA	EMTB	AVIA	AVIB
Begrijpend Leestest 4(Aarmoutse) ABL4	1,000	0,652	0,655	0,626	0,452	0,452	0,501	0,492
Begrijpend Lezen M3 (Cito) CBM3	0,652	1,000	0,622	0,631	0,508	0,516	0,526	0,531
Woordenschattest 4 (Aarmoutse) AWS4	0,655	0,622	1,000	0,752	0,452	0,458	0,475	0,475
Woordenschattest 5 (Aarmoutse) AWS5	0,626	0,631	0,752	1,000	0,424	0,427	0,445	0,458
Emt versie A EMTA	0,452	0,508	0,452	0,424	1,000	0,955	0,883	0,889
Emt versie B EMTB	0,452	0,516	0,458	0,427	0,955	1,000	0,892	0,891
AVI versie A AVIA	0,501	0,526	0,475	0,445	0,883	0,892	1,000	0,905
AVI versie B AVIB	0,492	0,531	0,475	0,458	0,889	0,891	0,905	1,000
Gemiddelde	27,0	16,9	27,3	20,6	45,0	43,4	4,9	5,0
Standaardafwijking	6,1	5,1	5,9	6,9	14,5	14,8	2,3	2,4

Tabel B19.5: Correlaties tussen geobserveerde variabelen en gemiddelden en standaardafwijkingen van de variabelen in cohort 4 periode 2 (najaar 1992; N=846)

	CBM3	ABL5	AWS5	SWS3	EMTA	EMTB	AVIA
Begrijpend Lezen M3 (Cito) CBM3	1,000	0,704	0,594	0,656	0,473	0,487	0,488
Begrijpend Leestest 5 (Aarmoutse) ABL5	0,704	1,000	0,703	0,712	0,482	0,492	0,519
Woordenschattest 5 (Aarmoutse) AWS5	0,594	0,703	1,000	0,759	0,437	0,440	0,464
Woordenschattest 3 (Stijnen) SWS3	0,656	0,712	0,759	1,000	0,523	0,518	0,538
Emt versie A EMTA	0,473	0,482	0,437	0,523	1,000	0,943	0,848
Emt versie B EMTB	0,487	0,492	0,440	0,518	0,943	1,000	0,852
AVI versie A AVIA	0,488	0,519	0,464	0,538	0,848	0,852	1,000
Gemiddelde	18,4	18,7	23,8	38,8	52,4	51,0	5,9
Standaardafwijking	5,4	6,6	6,9	8,6	14,8	14,8	2,3

Tabel B19.6: Correlaties tussen geobserveerde variabelen en gemiddelden en standaardafwijkingen van de variabelen in cohort 4 periode 3 (voorjaar 1993; N=843)

	ABL5	CBM4	SWS3	AWS6	EMTA	EMTB	AVIA	AVIB
Begrijpend Leestest 5(Aarmoutse) ABL5	1,000	0,725	0,671	0,559	0,415	0,429	0,457	0,459
Begrijpend Lezen M4 (Cito) CBM4	0,725	1,000	0,652	0,586	0,417	0,434	0,472	0,467
Woordenschattest 3 (Stijnen) SWS3	0,671	0,652	1,000	0,649	0,432	0,448	0,498	0,489
Woordenschattest 6 (Aarmoutse) AWS6	0,559	0,586	0,649	1,000	0,337	0,359	0,384	0,378
Emt versie A EMTA	0,415	0,417	0,432	0,337	1,000	0,944	0,814	0,813
Emt versie B EMTB	0,429	0,434	0,448	0,359	0,944	1,000	0,812	0,810
AVI versie A AVIA	0,457	0,472	0,498	0,384	0,814	0,812	1,000	0,852
AVI versie B AVIB	0,459	0,467	0,489	0,378	0,813	0,810	0,852	1,000
Gemiddelde	22,1	17,3	43,7	20,1	59,2	58,5	7,0	6,9
Standaardafwijking	5,7	5,2	7,6	5,1	13,8	14,2	2,1	2,2

Tabel B19.7: Correlaties tussen geobserveerde variabelen en gemiddelden en standaardafwijkingen van de variabelen in cohort 5 periode 1 (voorjaar 1992; N=852)

	ABL5	CBM4	SWS3	AWS6	EMTA	EMTB	AVIA	AVIB
Begrijpend Leestest 5(Aarmoutse) ABL5	1,000	0,723	0,584	0,614	0,364	0,361	0,440	0,438
Begrijpend Lezen M4 (Cito) CBM4	0,723	1,000	0,584	0,611	0,352	0,364	0,436	0,445
Woordenschattest 3 (Stijnen) SWS3	0,584	0,584	1,000	0,510	0,313	0,309	0,381	0,386
Woordenschattest 6 (Aarmoutse) AWS6	0,614	0,611	0,510	1,000	0,304	0,314	0,334	0,325
Emt versie A EMTA	0,364	0,352	0,313	0,304	1,000	0,936	0,822	0,809
Emt versie B EMTB	0,361	0,364	0,309	0,314	0,936	1,000	0,820	0,813
AVI versie A AVIA	0,440	0,436	0,381	0,334	0,822	0,820	1,000	0,880
AVI versie B AVIB	0,438	0,445	0,386	0,326	0,809	0,813	0,880	1,000
Gemiddelde	22,6	17,7	43,3	19,5	58,9	57,9	6,9	6,9
Standaardafwijking	5,2	4,6	8,1	5,2	13,5	13,4	1,9	2,0

Tabel B19.8: Correlaties tussen geobserveerde variabelen en gemiddelden en standaardafwijkingen van de variabelen in cohort 5 periode 2 (najaar 1992; N=806)

	CBM4	ABL6	AWS6	SWS4	EMTA	EMTB	AVIA
Begrijpend Lezen M4 (Cito) CBM4	1,000	0,682	0,562	0,673	0,385	0,390	0,427
Begrijpend Leestest 6(Aarmoutse) ABL6	0,682	1,000	0,628	0,658	0,453	0,444	0,456
Woordenschattest 6 (Aarmoutse) AWS6	0,562	0,628	1,000	0,731	0,395	0,386	0,348
Woordenschattest 4 (Stijnen) SWS4	0,673	0,658	0,731	1,000	0,459	0,457	0,440
Emt versie A EMTA	0,385	0,453	0,395	0,459	1,000	0,928	0,778
Emt versie B EMTB	0,390	0,444	0,386	0,457	0,928	1,000	0,776
AVI versie A AVIA	0,427	0,456	0,348	0,440	0,778	0,776	1,000
Gemiddelde	19,3	20,0	21,4	40,1	63,9	62,8	7,6
Standaardafwijking	4,6	5,4	5,6	9,1	13,6	13,7	1,9

Tabel B19.9: Correlaties tussen geobserveerde variabelen en gemiddelden en standaardafwijkingen van de variabelen in cohort 5 periode 3 (voorjaar 1993; N=802)

	ABL6	CBE4	SWS4	SWS5	EMTA	EMTB	AVIA	AVIB
Begrijpend Leestest 6(Aarmoutse) ABL6	1,000	0,710	0,696	0,695	0,372	0,360	0,385	0,420
Begrijpend Lezen E4 (Cito) CBE4	0,710	1,000	0,679	0,686	0,379	0,373	0,381	0,418
Woordenschattest 4 (Stijnen) SWS4	0,696	0,679	1,000	0,845	0,431	0,428	0,396	0,413
Woordenschattest 5 (Stijnen) SWS5	0,695	0,686	0,845	1,000	0,428	0,418	0,374	0,390
Emt versie A EMTA	0,372	0,379	0,431	0,428	1,000	0,936	0,752	0,741
Emt versie B EMTB	0,360	0,373	0,428	0,418	0,936	1,000	0,746	0,736
AVI versie A AVIA	0,385	0,381	0,396	0,374	0,752	0,746	1,000	0,858
AVI versie B AVIB	0,420	0,418	0,413	0,390	0,741	0,736	0,858	1,000
Gemiddelde	22,9	17,5	44,6	38,9	69,1	68,3	8,3	8,2
Standaardafwijking	5,2	4,3	8,1	8,5	13,4	13,7	1,7	1,8

Tabel B19.10: Correlaties tussen geobserveerde variabelen en gemiddelden en standaardafwijkingen van de variabelen in het longitudinale bestand van cohort 3 (N=859)

	VOORJAAR 1992 (groep 3)							
	CLB1	CLB2	BEGW	AWS3	EMTA	EMTB	AVIA	AVIB
VOORJAAR 1992 (groep 3)								
Lees en Begrijp 1 (Cito) CLB1	1,000	0,653	0,608	0,634	0,465	0,458	0,390	0,400
Lees en Begrijp 2 (Cito) CLB2	0,653	1,000	0,600	0,674	0,523	0,518	0,443	0,470
Begrijppentest Woorden(Aarmoutse) BEGW	0,608	0,600	1,000	0,657	0,454	0,438	0,407	0,415
Woordenschattest 3 (Aarmoutse) AWS3	0,634	0,674	0,657	1,000	0,461	0,469	0,420	0,439
Emt versie A EMTA	0,465	0,523	0,454	0,461	1,000	0,957	0,788	0,793
Emt versie B EMTB	0,458	0,518	0,438	0,469	0,957	1,000	0,795	0,803
AVI versie A AVIA	0,390	0,443	0,407	0,420	0,788	0,795	1,000	0,894
AVI versie B AVIB	0,400	0,470	0,415	0,439	0,793	0,803	0,894	1,000
NAJAAR 1992 (groep 4)								
Lees en Begrijp 2 (Cito) CLB2	0,589	0,690	0,517	0,548	0,483	0,466	0,439	0,445
Begrijpend Leestest 4 (Aarmoutse) ABL4	0,589	0,666	0,545	0,616	0,531	0,541	0,497	0,496
Woordenschattest 3 (Aarmoutse) AWS3	0,595	0,576	0,510	0,617	0,482	0,474	0,428	0,440
Woordenschattest 4 (Aarmoutse) AWS4	0,592	0,659	0,566	0,619	0,563	0,548	0,523	0,528
Emt versie A EMTA	0,501	0,571	0,445	0,484	0,828	0,827	0,758	0,774
Emt versie B EMTB	0,489	0,572	0,442	0,484	0,830	0,838	0,761	0,776
AVI versie A AVIA	0,473	0,514	0,412	0,471	0,731	0,732	0,739	0,743
VOORJAAR 1993 (groep 4)								
Begrijpend Lezen M3 (Cito) CBM3	0,499	0,562	0,465	0,505	0,400	0,398	0,368	0,380
Begrijpend Leestest 4(Aarmoutse) ABL4	0,580	0,574	0,521	0,604	0,406	0,405	0,353	0,360
Woordenschattest 4 (Aarmoutse) AWS4	0,557	0,607	0,543	0,618	0,455	0,443	0,417	0,428
Woordenschattest 5 (Aarmoutse) AWS5	0,483	0,535	0,511	0,569	0,436	0,428	0,419	0,427
Emt versie A EMTA	0,455	0,508	0,373	0,445	0,778	0,778	0,697	0,705
Emt versie B EMTB	0,461	0,512	0,382	0,447	0,766	0,766	0,684	0,690
AVI versie A AVIA	0,438	0,485	0,385	0,436	0,683	0,683	0,665	0,679
AVI versie B AVIB	0,449	0,490	0,380	0,441	0,701	0,695	0,671	0,693
Gemiddelde	22,7	21,3	28,8	19,3	26,2	24,7	1,2	1,2
Standaardafwijking	4,0	7,4	6,1	4,8	11,9	11,9	1,4	1,4

	NAJAAR 1992 (groep 4)						
	CLB2	ABL4	AWS3	AWS4	EMTA	EMTB	AVIA
VOORJAAR 1992 (groep 3)							
Lees en Begrijp 1 (Cito) CLB1	0,589	0,589	0,595	0,592	0,501	0,489	0,473
Lees en Begrijp 2 (Cito) CLB2	0,690	0,666	0,576	0,659	0,571	0,572	0,514
Begrippentest Woorden(Aamoutse) BEGW	0,517	0,545	0,510	0,566	0,445	0,442	0,412
Woordenschattest 3 (Aamoutse) AWS3	0,548	0,616	0,617	0,619	0,484	0,484	0,471
Emt versie A EMTA	0,483	0,531	0,482	0,563	0,828	0,830	0,731
Emt versie B EMTB	0,466	0,541	0,474	0,548	0,827	0,838	0,732
AVI versie A AVIA	0,439	0,497	0,428	0,523	0,758	0,761	0,739
AVI versie B AVIB	0,445	0,496	0,440	0,528	0,774	0,776	0,743
NAJAAR 1992 (groep 4)							
Lees en Begrijp 2 (Cito) CLB2	1,000	0,750	0,688	0,721	0,603	0,587	0,569
Begrijpend Leestest 4 (Aamoutse) ABL4	0,750	1,000	0,708	0,771	0,616	0,619	0,611
Woordenschattest 3 (Aamoutse) AWS3	0,688	0,708	1,000	0,709	0,579	0,568	0,558
Woordenschattest 4 (Aamoutse) AWS4	0,721	0,771	0,709	1,000	0,626	0,619	0,612
Emt versie A EMTA	0,603	0,616	0,579	0,626	1,000	0,968	0,848
Emt versie B EMTB	0,587	0,619	0,568	0,619	0,968	1,000	0,844
AVI versie A AVIA	0,569	0,611	0,558	0,612	0,848	0,844	1,000
VOORJAAR 1993 (groep 4)							
Begrijpend Lezen M3 (Cito) CBM3	0,603	0,645	0,538	0,627	0,486	0,477	0,468
Begrijpend Leestest 4(Aamoutse) ABL4	0,627	0,695	0,654	0,667	0,458	0,456	0,447
Woordenschattest 4 (Aamoutse) AWS4	0,681	0,697	0,686	0,756	0,516	0,505	0,513
Woordenschattest 5 (Aamoutse) AWS5	0,582	0,687	0,589	0,738	0,464	0,462	0,464
Emt versie A EMTA	0,552	0,590	0,552	0,583	0,908	0,909	0,808
Emt versie B EMTB	0,560	0,595	0,559	0,595	0,899	0,905	0,814
AVI versie A AVIA	0,545	0,568	0,519	0,574	0,790	0,786	0,809
AVI versie B AVIB	0,558	0,588	0,536	0,594	0,805	0,801	0,816
Gemiddelde	22,5	21,2	21,4	21,5	31,7	30,5	3,0
Standaardafwijking	6,9	7,5	4,3	7,4	13,7	14,1	2,0

	VOORJAAR 1993 (groep 4)							
	CBM3	ABL4	AWS4	AWS5	EMTA	EMTB	AVIA	AVIB
VOORJAAR 1992 (groep 3)								
Lees en Begrijp 1 (Cito) CLB1	0,499	0,580	0,557	0,483	0,455	0,461	0,438	0,449
Lees en Begrijp 2 (Cito) CLB2	0,562	0,574	0,607	0,535	0,508	0,512	0,485	0,490
Begrijppentest Woorden(Aarmoutse) BEGW	0,465	0,521	0,543	0,511	0,373	0,382	0,385	0,380
Woordenschattest 3 (Aarmoutse) AWS3	0,505	0,604	0,618	0,569	0,445	0,447	0,436	0,441
Emt versie A EMTA	0,400	0,406	0,455	0,436	0,778	0,766	0,683	0,701
Emt versie B EMTB	0,398	0,405	0,443	0,428	0,778	0,766	0,683	0,695
AVI versie A AVIA	0,368	0,353	0,417	0,419	0,697	0,684	0,665	0,671
AVI versie B AVIB	0,380	0,360	0,428	0,427	0,705	0,690	0,679	0,693
NAJAAR 1992 (groep 4)								
Lees en Begrijp 2 (Cito) CLB2	0,603	0,627	0,681	0,582	0,552	0,560	0,545	0,558
Begrijpend Leestest 4 (Aarmoutse) ABL4	0,645	0,695	0,697	0,687	0,590	0,595	0,568	0,588
Woordenschattest 3 (Aarmoutse) AWS3	0,538	0,654	0,686	0,589	0,552	0,559	0,519	0,536
Woordenschattest 4 (Aarmoutse) AWS4	0,627	0,667	0,756	0,738	0,583	0,595	0,574	0,594
Emt versie A EMTA	0,486	0,458	0,516	0,464	0,908	0,899	0,790	0,805
Emt versie B EMTB	0,477	0,456	0,505	0,462	0,909	0,905	0,786	0,801
AVI versie A AVIA	0,468	0,447	0,513	0,464	0,808	0,814	0,809	0,816
VOORJAAR 1993 (groep 4)								
Begrijpend Lezen M3 (Cito) CBM3	1,000	0,661	0,635	0,624	0,494	0,495	0,470	0,487
Begrijpend Leestest 4(Aarmoutse) ABL4	0,661	1,000	0,729	0,647	0,471	0,474	0,470	0,478
Woordenschattest 4 (Aarmoutse) AWS4	0,635	0,729	1,000	0,752	0,510	0,518	0,507	0,511
Woordenschattest 5 (Aarmoutse) AWS5	0,624	0,647	0,752	1,000	0,435	0,440	0,454	0,479
Emt versie A EMTA	0,494	0,471	0,510	0,435	1,000	0,963	0,796	0,817
Emt versie B EMTB	0,495	0,474	0,518	0,440	0,963	1,000	0,803	0,818
AVI versie A AVIA	0,470	0,470	0,507	0,454	0,796	0,803	1,000	0,857
AVI versie B AVIB	0,487	0,478	0,511	0,479	0,817	0,818	0,857	1,000
Gemiddelde	16,5	26,7	27,1	21,0	43,1	42,1	4,5	4,5
Standaardafwijking	5,4	6,3	6,1	6,8	14,0	14,4	2,1	2,1

Tabel B19.11: Correlaties tussen geobserveerde variabelen en gemiddelden en standaardafwijkingen van de variabelen in het longitudinale bestand van cohort 4 (N=796)

	VOORJAAR 1992 (groep 4)							
	ABL4	CBM3	AWS4	AWS5	EMTA	EMTB	AVIA	AVIB
VOORJAAR 1992 (groep 4)								
Begrijpend Leestest 4(Aarmoutse) ABL4	1,000	0,647	0,660	0,624	0,432	0,435	0,489	0,471
Begrijpend Lezen M3 (Cito) CBM3	0,647	1,000	0,629	0,641	0,508	0,513	0,529	0,528
Woordenschattest 4 (Aarmoutse) AWS4	0,660	0,629	1,000	0,749	0,441	0,447	0,475	0,470
Woordenschattest 5 (Aarmoutse) AWS5	0,624	0,641	0,749	1,000	0,432	0,433	0,458	0,467
Emt versie A EMTA	0,432	0,508	0,441	0,432	1,000	0,955	0,879	0,883
Emt versie B EMTB	0,435	0,513	0,447	0,433	0,955	1,000	0,886	0,885
AVI versie A AVIA	0,489	0,529	0,475	0,458	0,879	0,886	1,000	0,901
AVI versie B AVIB	0,471	0,528	0,470	0,467	0,883	0,885	0,901	1,000
NAJAAR 1992 (groep 5)								
Begrijpend Lezen M3 (Cito) CBM3	0,628	0,669	0,580	0,568	0,457	0,449	0,474	0,487
Begrijpend Leestest 5 (Aarmoutse) ABL5	0,660	0,653	0,654	0,681	0,470	0,475	0,516	0,504
Woordenschattest 5 (Aarmoutse) AWS5	0,644	0,637	0,736	0,838	0,435	0,439	0,474	0,470
Woordenschattest 3 (Stijnen) SWS3	0,695	0,638	0,741	0,731	0,501	0,511	0,543	0,543
Emt versie A EMTA	0,455	0,496	0,438	0,421	0,880	0,887	0,847	0,861
Emt versie B EMTB	0,449	0,503	0,445	0,427	0,881	0,901	0,855	0,864
AVI versie A AVIA	0,465	0,518	0,464	0,449	0,817	0,826	0,855	0,851
VOORJAAR 1993 (groep 5)								
Begrijpend Leestest 5(Aarmoutse) ABL5	0,665	0,581	0,589	0,592	0,403	0,399	0,441	0,447
Begrijpend Lezen M4 (Cito) CBM4	0,629	0,596	0,574	0,549	0,384	0,386	0,423	0,427
Woordenschattest 3 (Stijnen) SWS3	0,682	0,598	0,672	0,653	0,442	0,452	0,488	0,475
Woordenschattest 6 (Aarmoutse) AWS6	0,566	0,558	0,635	0,714	0,355	0,345	0,377	0,375
Emt versie A EMTA	0,430	0,477	0,408	0,395	0,862	0,869	0,833	0,840
Emt versie B EMTB	0,453	0,493	0,433	0,409	0,857	0,873	0,835	0,837
AVI versie A AVIA	0,462	0,524	0,474	0,425	0,780	0,790	0,814	0,810
AVI versie B AVIB	0,483	0,539	0,457	0,429	0,784	0,790	0,824	0,826
Gemiddelde	27,2	17,0	27,5	20,8	45,5	44,1	5,0	5,1
Standaardafwijking	6,0	5,1	5,8	7,0	14,3	14,5	2,2	2,3

	NAJAAR 1992 (groep 5)						
	CBM3	ABL5	AWS5	SWS3	EMTA	EMTB	AVIA
VOORJAAR 1992 (groep 4)							
Begrijpend Leestest 4(Aarmoutse) ABL4	0,628	0,660	0,644	0,695	0,455	0,449	0,465
Begrijpend Lezen M3 (Cito) CBM3	0,669	0,653	0,637	0,638	0,496	0,503	0,518
Woordenschattest 4 (Aarmoutse) AWS4	0,580	0,654	0,736	0,741	0,438	0,445	0,464
Woordenschattest 5 (Aarmoutse) AWS5	0,568	0,681	0,838	0,731	0,421	0,427	0,449
Emt versie A EMTA	0,457	0,470	0,435	0,501	0,880	0,881	0,817
Emt versie B EMTB	0,449	0,475	0,439	0,511	0,887	0,901	0,826
AVI versie A AVIA	0,474	0,516	0,474	0,543	0,847	0,855	0,855
AVI versie B AVIB	0,487	0,504	0,470	0,543	0,861	0,864	0,851
NAJAAR 1992 (groep 5)							
Begrijpend Lezen M3 (Cito) CBM3	1,000	0,699	0,597	0,652	0,466	0,484	0,486
Begrijpend Leestest 5 (Aarmoutse) ABL5	0,699	1,000	0,705	0,710	0,478	0,492	0,521
Woordenschattest 5 (Aarmoutse) AWS5	0,597	0,705	1,000	0,762	0,440	0,447	0,468
Woordenschattest 3 (Stijnen) SWS3	0,652	0,710	0,762	1,000	0,524	0,525	0,543
Emt versie A EMTA	0,466	0,478	0,440	0,524	1,000	0,943	0,847
Emt versie B EMTB	0,484	0,492	0,447	0,525	0,943	1,000	0,853
AVI versie A AVIA	0,486	0,521	0,468	0,543	0,847	0,853	1,000
VOORJAAR 1993 (groep 5)							
Begrijpend Leestest 5(Aarmoutse) ABL5	0,623	0,677	0,646	0,648	0,436	0,439	0,442
Begrijpend Lezen M4 (Cito) CBM4	0,617	0,637	0,603	0,613	0,426	0,426	0,440
Woordenschattest 3 (Stijnen) SWS3	0,613	0,658	0,702	0,743	0,450	0,454	0,488
Woordenschattest 6 (Aarmoutse) AWS6	0,500	0,605	0,744	0,657	0,354	0,363	0,380
Emt versie A EMTA	0,447	0,453	0,420	0,493	0,911	0,909	0,843
Emt versie B EMTB	0,459	0,467	0,437	0,522	0,901	0,922	0,841
AVI versie A AVIA	0,491	0,502	0,454	0,544	0,816	0,819	0,845
AVI versie B AVIB	0,507	0,506	0,456	0,533	0,809	0,819	0,848
Gemiddelde	18,4	18,7	23,8	38,9	52,4	51,1	6,00
Standaardafwijking	5,4	6,6	6,9	8,6	14,6	14,7	2,3

	VOORJAAR 1993 (groep 5)							
	ABL5	CBM4	SWS3	AWS6	EMTA	EMTB	AVIA	AVIB
VOORJAAR 1992 (groep 4)								
Begrijpend Leestest 4(Aarmoutse) ABL4	0,665	0,629	0,682	0,566	0,430	0,453	0,462	0,483
Begrijpend Lezen M3 (Cito) CBM3	0,581	0,596	0,598	0,558	0,477	0,493	0,524	0,539
Woordenschattest 4 (Aarmoutse) AWS4	0,589	0,574	0,672	0,635	0,408	0,433	0,474	0,457
Woordenschattest 5 (Aarmoutse) AWS5	0,592	0,549	0,653	0,714	0,395	0,409	0,425	0,429
Emt versie A EMTA	0,403	0,384	0,442	0,355	0,862	0,857	0,780	0,784
Emt versie B EMTB	0,399	0,386	0,452	0,345	0,869	0,873	0,790	0,790
AVI versie A AVIA	0,441	0,423	0,488	0,377	0,833	0,835	0,814	0,824
AVI versie B AVIB	0,447	0,427	0,475	0,375	0,840	0,837	0,810	0,826
NAJAAR 1992 (groep 5)								
Begrijpend Lezen M3 (Cito) CBM3	0,623	0,617	0,613	0,500	0,447	0,459	0,491	0,507
Begrijpend Leestest 5 (Aarmoutse) ABL5	0,677	0,637	0,658	0,605	0,453	0,467	0,502	0,506
Woordenschattest 5 (Aarmoutse) AWS5	0,646	0,603	0,702	0,744	0,420	0,437	0,454	0,456
Woordenschattest 3 (Stijnen) SWS3	0,648	0,613	0,743	0,657	0,493	0,522	0,544	0,533
Emt versie A EMTA	0,436	0,426	0,450	0,354	0,911	0,901	0,816	0,809
Emt versie B EMTB	0,439	0,426	0,454	0,363	0,909	0,922	0,819	0,819
AVI versie A AVIA	0,442	0,440	0,488	0,380	0,843	0,841	0,845	0,848
VOORJAAR 1993 (groep 5)								
Begrijpend Leestest 5(Aarmoutse) ABL5	1,000	0,722	0,669	0,561	0,410	0,432	0,459	0,460
Begrijpend Lezen M4 (Cito) CBM4	0,722	1,000	0,646	0,587	0,404	0,427	0,470	0,466
Woordenschattest 3 (Stijnen) SWS3	0,669	0,646	1,000	0,649	0,428	0,454	0,503	0,493
Woordenschattest 6 (Aarmoutse) AWS6	0,561	0,587	0,649	1,000	0,339	0,363	0,390	0,384
Emt versie A EMTA	0,410	0,404	0,428	0,339	1,000	0,943	0,812	0,812
Emt versie B EMTB	0,432	0,427	0,454	0,363	0,943	1,000	0,812	0,811
AVI versie A AVIA	0,459	0,470	0,503	0,390	0,812	0,812	1,000	0,850
AVI versie B AVIB	0,460	0,466	0,493	0,384	0,812	0,811	0,850	1,000
Gemiddelde	22,2	17,3	43,7	20,1	59,3	58,6	7,1	7,0
Standaardafwijking	5,6	5,1	7,5	5,1	13,7	14,1	2,1	2,2

Tabel B19.12: Correlaties tussen geobserveerde variabelen en gemiddelden en standaardafwijkingen van de variabelen in het longitudinale bestand van cohort 5 (N=739)

	VOORJAAR 1992 (groep 5)							
	ABL5	CBM4	SWS3	AWS6	EMTA	EMTB	AVIA	AVIB
VOORJAAR 1992 (groep 5)								
Begrijpend Leestest 5(Aarmoutse) ABL5	1,000	0,726	0,592	0,612	0,363	0,365	0,444	0,443
Begrijpend Lezen M4 (Cito) CBM4	0,726	1,000	0,587	0,620	0,354	0,371	0,442	0,452
Woordenschattest 3 (Stijnen) SWS3	0,592	0,587	1,000	0,497	0,322	0,321	0,392	0,403
Woordenschattest 6 (Aarmoutse) AWS6	0,612	0,620	0,497	1,000	0,312	0,322	0,344	0,342
Emt versie A EMTA	0,363	0,354	0,322	0,312	1,000	0,933	0,819	0,808
Emt versie B EMTB	0,365	0,371	0,321	0,322	0,933	1,000	0,816	0,809
AVI versie A AVIA	0,444	0,442	0,392	0,344	0,819	0,816	1,000	0,886
AVI versie B AVIB	0,443	0,452	0,403	0,342	0,808	0,809	0,886	1,000
NAJAAR 1992 (groep 6)								
Begrijpend Lezen M4 (Cito) CBM4	0,717	0,732	0,562	0,565	0,342	0,343	0,426	0,443
Begrijpend Leestest 6(Aarmoutse) ABL6	0,700	0,670	0,536	0,609	0,401	0,404	0,456	0,474
Woordenschattest 6 (Aarmoutse) AWS6	0,612	0,588	0,537	0,789	0,363	0,379	0,411	0,400
Woordenschattest 4 (Stijnen) SWS4	0,698	0,673	0,640	0,718	0,429	0,445	0,479	0,490
Emt versie A EMTA	0,406	0,387	0,354	0,342	0,905	0,891	0,821	0,807
Emt versie B EMTB	0,398	0,391	0,353	0,336	0,878	0,885	0,811	0,802
AVI versie A AVIA	0,432	0,434	0,395	0,320	0,765	0,773	0,824	0,825
VOORJAAR 1993 (groep 6)								
Begrijpend Leestest 6(Aarmoutse) ABL6	0,706	0,662	0,581	0,577	0,332	0,336	0,415	0,410
Begrijpend Lezen E4 (Cito) CBE4	0,664	0,649	0,514	0,533	0,355	0,358	0,423	0,415
Woordenschattest 4 (Stijnen) SWS4	0,714	0,669	0,672	0,694	0,392	0,400	0,451	0,453
Woordenschattest 5 (Stijnen) SWS5	0,685	0,655	0,595	0,703	0,406	0,410	0,437	0,441
Emt versie A EMTA	0,404	0,398	0,366	0,366	0,869	0,870	0,801	0,784
Emt versie B EMTB	0,390	0,395	0,365	0,348	0,855	0,873	0,789	0,780
AVI versie A AVIA	0,405	0,415	0,368	0,304	0,730	0,758	0,783	0,776
AVI versie B AVIB	0,418	0,445	0,388	0,319	0,720	0,744	0,784	0,780
Gemiddelde	22,7	17,7	43,3	19,4	59,0	58,1	7,0	7,0
Standaardafwijking	5,3	4,6	8,1	5,2	13,4	13,2	1,9	2,0

	NAJAAR 1992 (groep 6)						
	CBM4	ABL6	AWS6	SWS4	EMTA	EMTB	AVIA
VOORJAAR 1992 (groep 5)							
Begrijpend Leestest 5(Aarmoutse) ABL5	0,717	0,700	0,612	0,698	0,406	0,398	0,432
Begrijpend Lezen M4 (Cito) CBM4	0,732	0,670	0,588	0,673	0,387	0,391	0,434
Woordenschattest 3 (Stijnen) SWS3	0,562	0,536	0,537	0,640	0,354	0,353	0,395
Woordenschattest 6 (Aarmoutse) AWS6	0,565	0,609	0,789	0,718	0,342	0,336	0,320
Emt versie A EMTA	0,342	0,401	0,363	0,429	0,905	0,878	0,765
Emt versie B EMTB	0,343	0,404	0,379	0,445	0,891	0,885	0,773
AVI versie A AVIA	0,426	0,456	0,411	0,479	0,821	0,811	0,824
AVI versie B AVIB	0,443	0,474	0,400	0,490	0,807	0,802	0,825
NAJAAR 1992 (groep 6)							
Begrijpend Lezen M4 (Cito) CBM4	1,000	0,690	0,576	0,684	0,379	0,380	0,420
Begrijpend Leestest 6(Aarmoutse) ABL6	0,690	1,000	0,650	0,675	0,449	0,441	0,467
Woordenschattest 6 (Aarmoutse) AWS6	0,576	0,650	1,000	0,738	0,405	0,398	0,368
Woordenschattest 4 (Stijnen) SWS4	0,684	0,675	0,738	1,000	0,473	0,463	0,462
Emt versie A EMTA	0,379	0,449	0,405	0,473	1,000	0,927	0,777
Emt versie B EMTB	0,380	0,441	0,398	0,463	0,927	1,000	0,776
AVI versie A AVIA	0,420	0,467	0,368	0,462	0,777	0,776	1,000
VOORJAAR 1993 (groep 6)							
Begrijpend Leestest 6(Aarmoutse) ABL6	0,685	0,729	0,593	0,662	0,375	0,372	0,417
Begrijpend Lezen E4 (Cito) CBE4	0,668	0,601	0,556	0,645	0,385	0,380	0,421
Woordenschattest 4 (Stijnen) SWS4	0,655	0,672	0,710	0,819	0,424	0,414	0,429
Woordenschattest 5 (Stijnen) SWS5	0,624	0,665	0,694	0,789	0,430	0,428	0,431
Emt versie A EMTA	0,374	0,457	0,409	0,468	0,912	0,895	0,746
Emt versie B EMTB	0,384	0,432	0,405	0,463	0,885	0,906	0,750
AVI versie A AVIA	0,388	0,422	0,344	0,419	0,754	0,750	0,822
AVI versie B AVIB	0,437	0,439	0,351	0,440	0,741	0,733	0,817
Gemiddelde	19,3	19,9	21,3	40,1	63,8	62,9	7,6
Standaardafwijking	4,7	5,4	5,6	9,0	13,4	13,5	1,9

	VOORJAAR 1993 (groep 6)							
	ABL6	CBE4	SWS4	SWS5	EMTA	EMTB	AVIA	AVIB
VOORJAAR 1992 (groep 5)								
Begrijpend Leestest 5(Aarmoutse) ABL5	0,706	0,664	0,714	0,685	0,404	0,390	0,405	0,418
Begrijpend Lezen M4 (Cito) CBM4	0,662	0,649	0,669	0,655	0,398	0,395	0,415	0,445
Woordenschattest 3 (Stijnen) SWS3	0,581	0,514	0,672	0,595	0,366	0,365	0,368	0,388
Woordenschattest 6 (Aarmoutse) AWS6	0,577	0,533	0,694	0,703	0,366	0,348	0,304	0,319
Emt versie A EMTA	0,332	0,355	0,392	0,406	0,869	0,855	0,730	0,720
Emt versie B EMTB	0,336	0,358	0,400	0,410	0,870	0,873	0,758	0,744
AVI versie A AVIA	0,415	0,423	0,451	0,437	0,801	0,789	0,783	0,784
AVI versie B AVIB	0,410	0,415	0,453	0,441	0,784	0,780	0,776	0,780
NAJAAR 1992 (groep 6)								
Begrijpend Lezen M4 (Cito) CBM4	0,685	0,668	0,655	0,624	0,374	0,384	0,388	0,437
Begrijpend Leestest 6(Aarmoutse) ABL6	0,729	0,601	0,672	0,665	0,457	0,432	0,422	0,439
Woordenschattest 6 (Aarmoutse) AWS6	0,593	0,556	0,710	0,694	0,409	0,405	0,344	0,351
Woordenschattest 4 (Stijnen) SWS4	0,662	0,645	0,819	0,789	0,468	0,463	0,419	0,440
Emt versie A EMTA	0,375	0,385	0,424	0,430	0,912	0,885	0,754	0,741
Emt versie B EMTB	0,372	0,380	0,414	0,428	0,895	0,906	0,750	0,733
AVI versie A AVIA	0,417	0,421	0,429	0,431	0,746	0,750	0,822	0,817
VOORJAAR 1993 (groep 6)								
Begrijpend Leestest 6(Aarmoutse) ABL6	1,000	0,706	0,705	0,704	0,371	0,358	0,378	0,413
Begrijpend Lezen E4 (Cito) CBE4	0,706	1,000	0,676	0,688	0,377	0,368	0,378	0,415
Woordenschattest 4 (Stijnen) SWS4	0,705	0,676	1,000	0,852	0,430	0,424	0,396	0,413
Woordenschattest 5 (Stijnen) SWS5	0,704	0,688	0,852	1,000	0,431	0,417	0,384	0,393
Emt versie A EMTA	0,371	0,377	0,430	0,431	1,000	0,935	0,752	0,743
Emt versie B EMTB	0,358	0,368	0,424	0,417	0,935	1,000	0,741	0,734
AVI versie A AVIA	0,378	0,378	0,396	0,384	0,752	0,741	1,000	0,856
AVI versie B AVIB	0,413	0,415	0,413	0,393	0,743	0,734	0,856	1,000
Gemiddelde	22,8	17,5	44,6	38,9	69,3	68,6	8,4	8,2
Standaardafwijking	5,2	4,3	8,0	8,5	13,3	13,5	1,7	1,8

CURRICULUM VITAE

Jacqueline Johanna Cornelia Maria Visser werd op 2 september 1962 geboren te Beek gemeente Bergh. In 1980 behaalde zij het VWO-diploma aan het St. Ludgercollege te Doetinchem. Aan de Katholieke Universiteit Nijmegen slaagde zij in 1983 cum laude voor het kandidaatsexamen psychologie. Daarna volgde zij de studie Interdisciplinaire Onderwijskunde aan de Katholieke Universiteit Nijmegen, met als hoofdrichting 'Vormgeving en begeleiding van onderwijsleerprocessen' en als bijvak mathematische psychologie. Op 23 februari 1987 slaagde zij cum laude voor het doctoraalexamen. Haar doctoraalscriptie werd onderscheiden door de faculteit der sociale wetenschappen wegens verricht wetenschappelijk werk.

Tussen november 1986 en januari 1991 was zij werkzaam als sociaal-wetenschappelijk onderzoeker bij het verkeerskundig adviesbureau Goudappel Coffeng te Deventer. Daar was ze betrokken bij een grootschalig longitudinaal mobiliteitsonderzoek. Sinds januari 1991 is zij verbonden aan de vakgroep onderzoek van KPC Groep te 's-Hertogenbosch. Daar is ze verantwoordelijk voor diverse interne en externe evaluatie- en marktonderzoeken, met een nadruk op kwantitatieve methoden en technieken van onderzoek. Een belangrijk onderzoek betrof de schaling, normering en validering van de nieuwe AVI-toetskaarten. Daarnaast verrichtte ze onderzoek op het terrein van leerlingbegeleiding, jeugdcriminaliteit en levensbeschouwelijk onderwijs en voerde ze diverse onderzoeken uit naar het imago en de leerlingstromen van scholen.



KPC Groep
Postbus 482
5201 AL 's-Hertogenbosch

1.029.61